

DISTRIBUTED LINEAR PARAMETER ESTIMATION: ASYMPTOTICALLY EFFICIENT ADAPTIVE STRATEGIES

SOUMMYA KAR^{†§}, JOSÉ M. F. MOURA^{†§}, AND H. VINCENT POOR^{‡¶}

Abstract. The paper considers the problem of distributed adaptive linear parameter estimation in multi-agent inference networks. Local sensing model information is only partially available at the agents and inter-agent communication is assumed to be unpredictable. The paper develops a generic mixed time-scale stochastic procedure consisting of simultaneous distributed learning and estimation, in which the agents adaptively assess their relative observation quality over time and fuse the innovations accordingly. Under rather weak assumptions on the statistical model and the inter-agent communication, it is shown that, by properly tuning the consensus potential with respect to the innovation potential, the asymptotic information rate loss incurred in the learning process may be made negligible. As such, it is shown that the agent estimates are asymptotically efficient, in that their asymptotic covariance coincides with that of a centralized estimator (the inverse of the centralized Fisher information rate for Gaussian systems) with perfect global model information and having access to all observations at all times. The proof techniques are mainly based on convergence arguments for non-Markovian mixed time scale stochastic approximation procedures. Several approximation results developed in the process are of independent interest.

Key words. Multi-Agent Systems, Distributed Estimation, Mixed time scale, Stochastic approximation, Asymptotically Efficient, Adaptive Algorithms.

1. Introduction.

1.1. Background and Motivation. Recent advances in sensing and communication technologies have enabled the proliferation of heterogeneous sensing resources in multi-agent networks, typical examples being cyberphysical systems and distributed sensor networks. Due to the large size of these networks and the presence of geographically spread resources, distributed information processing and optimization (see, for example, [1, 2]) techniques are gaining prominence. They not only offer a robust alternative to fusion center based centralized approaches, but lead to efficient use of the network resources by distributing the computing and communication burden among the agents. A key challenge in such distributed processing involves the lack of global (sensing) model information at the local agent level. Moreover, the systems under consideration are dynamic, often leading to uncertainty in the spatial distribution of the information content. The performance of existing distributed information processing and optimization schemes (see, for example, [3–17]) based on accurate knowledge of the sensed data statistics may suffer substantially in the face of such parametric uncertainties. This necessitates the development of adaptive schemes that learn the model parameters over time in conjunction with carrying out the desired information processing task.

Motivated by the above, in this paper we focus on the problem of distributed recursive least squares parameter estimation, in which the agents have no prior knowledge of the global sensing model and of the individual observation qualities as measured in terms of the signal to noise ratio (SNR). Our goal is to develop an adaptive

[†]Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA (soumyak@andrew.cmu.edu, moura@ece.cmu.edu).

[‡]Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA (poor@princeton.edu).

[§]The work of S. Kar and J.M.F. Moura was partially supported by NSF grants # CCF-1018509 and CCF-1011903 and by AFOSR grant # FA9550101291.

[¶]The work of H. Vincent Poor was supported by the Office of Naval Research under grant # N00014-12-1-0767.

distributed scheme that is asymptotically efficient, i.e., achieves the same estimation performance at each agent (in terms of asymptotic covariance) as that of a (hypothetical) centralized fusion center with perfect global model information and having access to all agents' observations at all times. To this end, we develop a *consensus+innovation* scheme, [6], in which the agents collaborate by exchanging (appropriate) messages with their neighbors (consensus) and fusing the acquired information with the new local observation (innovation). Apart from the issue of optimality, the inter-agent collaboration is necessary for estimator consistency, as the local observations are generally not rich enough to guarantee global observability. Lacking prior global model and local SNR information, the innovation gains at the agents are not optimal *a priori*, and the agents simultaneously engage in a distributed learning process based on past data samples with the aim of recovering the optimal gains asymptotically. Thus the distributed learning process proceeds in conjunction and interacts with the estimate update. Intuitively, the overall update scheme has the structure of a certainty-equivalent control system (see, for example, [18, 19] and the references therein, in the context of parameter estimation), the key difference being the distributed nature of the learning and estimation tasks. Under rather weak assumptions on the inter-agent communication (network connectivity on *average*) we show that, by properly tuning the consensus potential with respect to the innovation potential, the asymptotic information rate loss incurred in the learning process may be made negligible, and the agent estimates are asymptotically efficient in that their asymptotic covariances coincide with that of the hypothetical centralized estimator. The proper tuning of the persistent consensus and innovation potentials are necessary for this optimality, leading to a mixed time-scale stochastic procedure. In this context, we note the study of mixed time-scale stochastic procedures that arise in algorithms of the simulated annealing type (see, for example, [20]). Apart from being distributed, our scheme technically differs from [20] in that, whereas the additive perturbation in [20] is a martingale difference sequence, ours is a network dependent consensus potential manifesting past dependence. In fact, intuitively, a key step in the analysis is to derive pathwise strong approximation results to characterize the rate at which the consensus term/process converges to a martingale difference process. We also emphasize that our notion of mixed time-scale is different from that of stochastic algorithms with coupling (see [21, 22]), where a quickly switching parameter influences the relatively slower dynamics of another state, leading to *averaged* dynamics. Mixed time scale procedures of this latter type arise in multi-scale distributed information diffusion problems; see, in particular, the paper [23] that studies interactive consensus formation in Markov modulated switching networks.

We comment on the main technical ingredients of the paper. Due to the mixed time-scale behavior and the non-Markovianity (induced by the learning process that uses all past information), the stochastic procedure does not fall under the purview of standard stochastic approximation (see, for example, [24]) or distributed stochastic approximation (see, for example, [6, 25–31]) procedures. As such, we develop several intermediate results on the pathwise convergence rates of mixed time-scale stochastic procedures. Some of these tools are of independent interest and general enough to be applicable to other distributed adaptive information processing problems.

We briefly summarize the organization of the rest of the paper. Section 1.2 presents notation to be used throughout. The abstract problem formulation and the mixed time-scale distributed estimation scheme are stated and discussed in Sections 2.1 and 2.2 respectively. The main results of the paper are stated in Section 3,

whereas Section 4 presents some intermediate convergence results on recursive stochastic schemes. The key technical ingredients concerning the asymptotic properties of the distributed learning and estimation processes are obtained in Section 5, while the main results of the paper are proved in Section 6. Finally, Section 7 concludes the paper.

1.2. Notation. We denote the k -dimensional Euclidean space by \mathbb{R}^k . The set of reals is denoted by \mathbb{R} , whereas \mathbb{R}_+ denotes the non-negative reals. For $a, b \in \mathbb{R}$, we will use the notations $a \vee b$ and $a \wedge b$ to denote the maximum and minimum of a and b respectively. The set of $k \times k$ real matrices is denoted by $\mathbb{R}^{k \times k}$. The corresponding subspace of symmetric matrices is denoted by \mathbb{S}^k . The cone of positive semidefinite matrices is denoted by \mathbb{S}_+^k , whereas \mathbb{S}_{++}^k denotes the subset of positive definite matrices. The $k \times k$ identity matrix is denoted by I_k , while $\mathbf{1}_k$ and $\mathbf{0}_k$ denote respectively the column vector of ones and zeros in \mathbb{R}^k . Often the symbol $\mathbf{0}$ is used to denote the $k \times p$ zero matrix, the dimensions being clear from the context. The operator $\|\cdot\|$ applied to a vector denotes the standard Euclidean \mathcal{L}_2 norm, while applied to matrices it denotes the induced \mathcal{L}_2 norm, which is equivalent to the matrix spectral radius for symmetric matrices. The notation $A \otimes B$ is used to denote the Kronecker product of two matrices A and B .

Time is assumed to be discrete or slotted throughout the paper. The symbols t and s denote time, and \mathbb{T}_+ is the discrete index set $\{0, 1, 2, \dots\}$. The parameter to be estimated belongs to a subset Θ (generally open) of the Euclidean space \mathbb{R}^M . The true (but unknown) value of the parameter is θ^* and a canonical element of Θ is θ . The estimate of θ^* at time t at agent n is $\mathbf{x}_n(t) \in \mathbb{R}^M$. Without loss of generality, the initial estimate, $\mathbf{x}_n(0)$, at time 0 at agent n is a non-random quantity.

Spectral graph theory: The inter-agent communication topology may be described by an *undirected* graph $G = (V, E)$, with $V = [1 \dots N]$ and E the set of agents (nodes) and communication links (edges), respectively. The unordered pair $(n, l) \in E$ if there exists an edge between nodes n and l . We consider simple graphs, i.e., graphs devoid of self-loops and multiple edges. A graph is connected if there exists a path¹, between each pair of nodes. The neighborhood of node n is

$$\Omega_n = \{l \in V \mid (n, l) \in E\}.$$

Node n has degree $d_n = |\Omega_n|$ (the number of edges with n as one end point.) The structure of the graph can be described by the symmetric $N \times N$ adjacency matrix, $A = [A_{nl}]$, $A_{nl} = 1$, if $(n, l) \in E$, $A_{nl} = 0$, otherwise. Let the degree matrix be the diagonal matrix $D = \text{diag}(d_1 \dots d_N)$. By definition, the positive semidefinite matrix $L = D - A$ is called the graph Laplacian matrix. The eigenvalues of L can be ordered as $0 = \lambda_1(L) \leq \lambda_2(L) \leq \dots \leq \lambda_N(L)$, the eigenvector corresponding to $\lambda_1(L)$ being $(1/\sqrt{N})\mathbf{1}_N$. The multiplicity of the zero eigenvalue equals the number of connected components of the network; for a connected graph, $\lambda_2(L) > 0$. This second eigenvalue is the algebraic connectivity or the Fiedler value of the network; see [32] for detailed treatment of graphs and their spectral theory.

2. Problem Formulation.

2.1. System Model and Preliminaries. Let $\theta^* \in \Theta$ be an M -dimensional (vector) parameter that is to be estimated by a network of N agents. Throughout,

¹A path between nodes n and l of length m is a sequence $(n = i_0, i_1, \dots, i_m = l)$ of vertices, such that $(i_k, i_{k+1}) \in E \forall 0 \leq k \leq m - 1$.

we assume that all the random objects are defined on a common measurable space (Ω, \mathcal{F}) equipped with a filtration $\{\mathcal{F}_t\}$. For the true (but unknown) parameter value θ^* , probability and expectation are denoted by $\mathbb{P}_{\theta^*}[\cdot]$ and $\mathbb{E}_{\theta^*}[\cdot]$, respectively. All inequalities involving random variables are to be interpreted a.s. (almost surely.)

Each agent makes i.i.d. (independent and identically distributed) observations of noisy linear functions of the parameter. The observation model for the n -th agent is

$$\mathbf{y}_n(t) = H_n \theta^* + \zeta_n(t)$$

where: i) $\{\mathbf{y}_n(t) \in \mathbb{R}^{M_n}\}$ is the observation sequence for the n -th agent; and ii) for each n , $\{\zeta_n(t)\}$ is a zero-mean temporally i.i.d. noise sequence of bounded variance, such that, $\zeta_n(t)$ is \mathcal{F}_{t+1} adapted and independent of \mathcal{F}_t . Moreover, the sequences $\{\zeta_n(t)\}$ and $\{\zeta_l(t)\}$ are mutually uncorrelated for $n \neq l$. For most practical agent network applications, each agent observes only a subset of M_n of the components of θ , with $M_n \ll M$. It is then necessary for the agents to collaborate by means of occasional local inter-agent message exchanges to achieve a reasonable estimate of the parameter θ^* . Moreover, due to inherent uncertainties in the deployment and the sensing environment, the statistics of the observation process (i.e., of the noise) are likely to be unknown *a priori*. For example, the exact observation noise variance at an agent depends on several factors beyond the control of the deployment process and should be learned over time for reasonable estimation performance. In other words, prior knowledge of the spatial distribution of the information content (e.g., which agent is more accurate than the others) may not be available, and the proposed estimation approach should be able to adaptively learn the true value of information leading to an accurate weighting of the various observation resources.

Let $R_n \in \mathbb{S}_{++}^{M_n}$ be the true covariance of the observation noise $\zeta_n(t)$ at agent n . It is well known that, given perfect knowledge of R_n for all n , the best linear centralized estimator $\{\mathbf{x}_c(t)\}$ of θ^* is asymptotically normal, i.e.,

$$\sqrt{t+1}(\mathbf{x}_c(t) - \theta^*) \implies \mathcal{N}(\mathbf{0}, \Sigma_c^{-1}),$$

provided the matrix $\Sigma_c = \sum_{n=1}^N H_n^T R_n^{-1} H_n$ is invertible. In case the observation process is Gaussian, the best linear estimator is optimal, and Σ_c coincides with the Fisher information rate. In general, with the knowledge of the covariance only and no other specifics about the noise distribution, the above estimate is optimal, in that no other estimate achieves smaller asymptotic covariance than Σ_c^{-1} for all distributions with covariance R_n .

The goal of this paper is to develop a distributed estimator that leads to asymptotically normal estimates with the same asymptotic covariance Σ_c^{-1} at each agent under the following constraints: (1) Each agent is aware only of its local observation model H_n and, more importantly, (2) the true noise covariance R_n is not known *a priori* at agent n and needs to be learned from the received observation samples and exchanged messages with its neighbors over time. Recently, in [33] a distributed algorithm was introduced that leads to the desired centralized asymptotic covariance at each agent but requires full model information (i.e., all the H_n 's) and the exact covariance values R_n at all agents. This is due to the fact that, for optimal asymptotic covariance, the approach in [33] requires an appropriate innovation gain at each agent, the latter depending on all the model matrices and noise covariances. In the absence of model and covariance information, a natural alternative is to employ a certainty-equivalence type approach in which an adaptive sequential gain refinement (learning)

step is incorporated into the desired estimation task. In this paper, we show that such a learning process (see Section 2.2) is feasible in a distributed setting and, more importantly, the coupling between the learning and parameter estimation tasks does not slow down the convergence rate (measured in terms of asymptotic covariance) of the latter to θ^* .

2.2. Adaptive Distributed Estimator: Algorithm \mathcal{ADLE} . The adaptive distributed linear estimator (\mathcal{ADLE}) involves two simultaneous update rules, namely, (1) the estimate (state) update and (2) the gain update. To formalize, let $\{\mathbf{x}_n(t)\}$ denote the $\{\mathcal{F}_t\}$ adapted sequence of estimates of θ^* at agent n .

Estimate Update: The estimate update at agent n then proceeds as follows:

$$\mathbf{x}_n(t+1) = \mathbf{x}_n(t) - \beta_t \sum_{l \in \Omega_n(t)} (\mathbf{x}_n(t) - \mathbf{x}_l(t)) + \alpha_t K_n(t) (\mathbf{y}_n(t) - H_n \mathbf{x}_n(t)). \quad (2.1)$$

In the above, $\{\beta_t\}$ and $\{\alpha_t\}$ represent appropriate time-varying weighting factors for the consensus (agreement) and innovation (new observation) potentials respectively, whereas $\{K_n(t)\}$ is an adaptively chosen $\{\mathcal{F}_t\}$ -adapted matrix gain process. Also, $\Omega_n(t)$ denotes the $\{\mathcal{F}_{t+1}\}$ -adapted time-varying random neighborhood of agent n at time t .

Gain Update: The adaptive gain update at sensor n involves another $\{\mathcal{F}_t\}$ adapted distributed learning process that proceeds in parallel with the estimate update. In particular, we set

$$K_n(t) = (G_n(t) + \gamma_t I_M)^{-1} H_n^T (Q_n(t) + \gamma_t I_{M_n})^{-1} \quad (2.2)$$

where $\{\gamma_t\}$ is a sequence of positive reals, such that $\gamma_t \rightarrow 0$ as $t \rightarrow \infty$, and the positive semidefinite matrix sequences $\{Q_n(t)\}$ and $\{G_n(t)\}$ evolve as follows:

$$Q_n(t+1) = \frac{1}{t} \sum_{s=0}^t \mathbf{y}_n(s) \mathbf{y}_n^T(s) - \left(\frac{1}{t} \sum_{s=0}^{t-1} \mathbf{y}_n(s) \right) \left(\frac{1}{t} \sum_{s=0}^{t-1} \mathbf{y}_n(s) \right)^T, \quad (2.3)$$

and

$$G_n(t+1) = G_n(t) - \beta_t \sum_{l \in \Omega_n(t)} (G_n(t) - G_l(t)) + \alpha_t \left(H_n^T (Q_n(t) + \gamma_t I_{M_n})^{-1} H_n - G_n(t) \right) \quad (2.4)$$

with positive semidefinite initial conditions $Q_n(0)$ and $G_n(0)$ respectively.

Before discussing further, we formalize assumptions on the model, the time-varying communication topology and the algorithm weight sequences $\{\alpha_t\}$ and $\{\beta_t\}$ in the following:

(A.1): *The observation model is globally observable, i.e., the (normalized) Gramian matrix*

$$\bar{\Sigma}_c = \frac{1}{N} \sum_{n=1}^N H_n^T R_n^{-1} H_n$$

is invertible, where R_n denotes the non-singular true (but unknown) covariance of the observation noise $\zeta_n(t)$ at agent n .

(A.2): *The $\{\mathcal{F}_{t+1}\}$ -adapted sequence $\{L_t\}$ of communication network Laplacians (modeling the agent communication neighborhoods $\{\Omega_n(t)\}$ at each time t) is temporally i.i.d. with L_t being independent of \mathcal{F}_t for each t . Further, the sequence $\{L_t\}$*

is connected on the average, i.e., $\lambda_2(\bar{L}) > 0$, where $\bar{L} = \mathbb{E}_{\theta^*}[L_t]$ denotes the mean Laplacian.

(A.3): The sequences $\{L_t\}$ and $\{\zeta_n(t)\}_{n \in V}$ are mutually independent.

(A.4): There exists $\varepsilon_1 > 0$, such that for all n , $\mathbb{E}_{\theta^*}[\|\zeta_n(t)\|^{2+\varepsilon_1}] < \infty$.

(A.5): The weight sequences $\{\alpha_t\}$ and $\{\beta_t\}$ are given by

$$\alpha_t = \frac{a}{(t+1)^{\tau_1}} \quad \text{and} \quad \beta_t = \frac{b}{(t+1)^{\tau_2}}, \quad (2.5)$$

where $a, b > 0$, $0 < \tau_2 \leq \tau_1 \leq 1$ and $\tau_1 > \tau_2 + 1/(2 + \varepsilon_1) + 1/2$.

REMARK 2.1. Note that the global observability requirement in **(A.1)** is quite weak and, in fact, is necessary to attain estimator consistency in a centralized setting. In a sense, the assumption **(A.1)** on the global sensing model and the *mean connectivity* condition in **(A.2)** provide minimal structural conditions for attaining *distributed observability*, i.e., the ability to obtain consistent parameter estimates in the proposed distributed information setting. Intuitively, the necessity of **(A.2)** (in addition to **(A.1)**) for such distributed observability stems from the observation that, in general, in the absence of local observability a disconnected inter-agent communication network would lead to multiple communication-disjoint agent components, none with sufficient informative measurements to consistently estimate the true parameter. We emphasize that the *mean* network connectivity assumption formalized in **(A.2)**, which generalizes the notion of connectivity in static communication topologies to dynamic stochastic scenarios, models a broad class of agent networks with unpredictable communication; for instance, **(A.2)** allows for spatially correlated communication link failures (often resulting from multi-agent interference) and subsumes the commonly used packet erasure model in gossip type of agent communications [34]. On the other hand, in the current setting, we assume that the inter-agent communication is noise-free and unquantized in the event of an active communication link; the problem of quantized data exchange in networked control systems (see, for example, [6, 35–37]) is an active research topic.

We comment on the choice of the weight sequences $\{\beta_t\}$ and $\{\alpha_t\}$ associated with the consensus and innovation potentials respectively. From **(A.5)** we note that both the excitations for agent-collaboration (consensus) and local innovation are persistent, i.e., the sequences $\{\beta_t\}$ and $\{\alpha_t\}$ sum to ∞ - a standard requirement in stochastic approximation type algorithms to drive the updates to the desired limit from arbitrary initial conditions. Further, the square summability of $\{\alpha_t\}$ ($\tau_1 > 1/2$) is required to mitigate the effect of stochastic sensing noise perturbing the innovations. The requirement $\beta_t/\alpha_t \rightarrow \infty$ as $t \rightarrow \infty$ ($\tau_1 > \tau_2$), i.e., the asymptotic domination of the consensus potential over the local innovations ensures the right information mixing thus, as shown below, leading to optimal estimation performance. Technically, the different asymptotic decay rates of the two potentials lead to mixed time-scale stochastic recursions whose analyses require new techniques in stochastic approximation as developed in the paper.

EXAMPLE 2.1. As an illustration, consider the agent model in Fig. 2.1 with $N = 5$ agents. The vector parameter $\theta^* \in \mathbb{R}^5$ in this example may have a physical interpretation, for example, with θ_n^* , the n -th component of θ^* , indicating the (unknown) intensity of a source geographically co-located with agent n , $n = 1, \dots, 5$. Each agent n observes a scalar sequence

$$y_n(t) = (\theta_{n-1}^* + \theta_n^* + \theta_{n+1}^*) + \zeta_n(t),$$

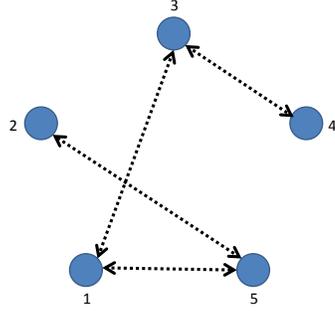


FIG. 2.1. An example: circles depict agents and dotted lines bi-directional communication links.

perhaps corresponding to a superposition of local source intensities, where we adopt the convention that $\theta_0^* = \theta_5^*$ and $\theta_6^* = \theta_1^*$. It is readily seen that the local agent observations are not globally observable for θ^* . In fact, in this example, without collaboration, no agent n would be able to reconstruct even the local intensity θ_n . The collective observation model is however globally observable for θ^* , i.e., **(A.1)** holds. The dotted lines denote the potential inter-agent communication links (possibly switching stochastically between on and off) through which the locally unobservable agents may collaborate by information exchange. By abstracting the above model in terms of the generic notation, the \mathcal{ADLE} estimate update rule (2.1) at an agent n , say $n = 3$, then takes the form

$$\begin{aligned} \mathbf{x}_3(t+1) &= \mathbf{x}_3(t) - \beta_t (2\mathbf{x}_3(t) - \mathbf{x}_1(t) - \mathbf{x}_4(t)) \\ &\quad + \alpha_t (G_3(t) + \gamma_t I_5)^{-1} H_3^T (Q_3(t) + \gamma_t)^{-1} (y_3(t) - x_{3,2}(t) - x_{3,3}(t) - x_{3,4}(t)), \end{aligned}$$

where $H_3 = [0 \ 1 \ 1 \ 1 \ 0]$, $Q_3(t)$ denotes the (scalar) sample covariance (2.3), $x_{3,l}(t)$ denotes the l -th component of $\mathbf{x}_3(t)$ with $l = 2, 3, 4$, and $G_3(t)$ is updated as

$$G_3(t+1) = G_3(t) - \beta_t (2G_3(t) - G_1(t) - G_4(t)) + \alpha_t \left(H_3^T (Q_3(t) + \gamma_t)^{-1} H_3 - G_3(t) \right).$$

In the above, we assumed that at time t , both the communication links (1,3) and (3,4) are active. Assuming that the stochasticity, if any, in the link formation satisfies **(A.2)**, the following analysis will show that the above estimate sequences will *optimally* converge to θ^* a.s. as $t \rightarrow \infty$.

3. Main Results. We formally state the main results of the paper, the proofs being provided in Section 6.

The first result concerns the asymptotic agreement or consensus among the various agent estimates.

THEOREM 3.1. *Let assumptions **(A.1)**-**(A.5)** hold. Then for each τ_0 such that*

$$0 \leq \tau_0 < \tau_1 - \tau_2 - \frac{1}{2 + \varepsilon_1},$$

we have

$$\mathbb{P}_{\theta^*} \left(\lim_{t \rightarrow \infty} (t+1)^{\tau_0} \|\mathbf{x}_n(t) - \mathbf{x}_l(t)\| = 0 \right) = 1$$

for any pair of agents n and l .

Theorem 3.1 relates the rate of agreement to the difference $\tau_1 - \tau_2$ of the algorithm weight parameters, the latter quantifying the relative intensities of the *global agreement* and *local innovation* potentials. Notably, the order of this convergence is independent of the network topology (as long as it is connected in the mean) and the distributed gain learning process (2.2)-(2.4). In fact, as will be evident from the proof arguments, the local covariance learning step in (2.3) may be replaced by any other consistent learning procedure, still retaining the order of convergence in Theorem 3.1.

THEOREM 3.2. *Let assumptions (A.1)-(A.5) hold with $\tau_1 = 1$ and $a \geq 1$. Then, for each n the estimate sequence $\{\mathbf{x}_n(t)\}$ is strongly consistent. In particular, we have*

$$\mathbb{P}_{\theta^*} \left(\lim_{t \rightarrow \infty} (t+1)^\tau \|\mathbf{x}_n(t) - \theta^*\| = 0 \right) = 1 \quad (3.1)$$

for each n and $\tau \in [0, 1/2)$.

The consistency in Theorem 3.2 is order optimal in that (3.1) fails to hold (unless the noise covariances are degenerate) with an exponent $\tau \geq 1/2$ for any (including centralized) estimation procedure, which is due to the fact that the optimal (centralized) estimator is asymptotically normal with non-degenerate (positive definite) asymptotic covariance.

The next result concerns the asymptotic efficiency of the estimates generated by the distributed \mathcal{ADLE} .

THEOREM 3.3. *Let assumptions (A.1)-(A.5) hold with $\tau_1 = 1$ and $a = 1$, and let $\Sigma_c = N\bar{\Sigma}_c$. Then, for each n*

$$\sqrt{t+1} (\mathbf{x}_n(t) - \theta^*) \implies \mathcal{N}(\mathbf{0}, \Sigma_c^{-1}),$$

where $\mathcal{N}(\cdot, \cdot)$ and \implies denote the Gaussian distribution and weak convergence, respectively.

Referring to the introductory discussion in Section 2.1, we note that the distributed and adaptive \mathcal{ADLE} achieves the best error covariance decay among the class of linear centralized estimators and is optimal in the Fisher information sense if the noise process is Gaussian. In a sense, Theorem 3.3 reinforces the applicability and advantage of distributed estimation schemes. Apart from issues of robustness, implementing a centralized estimator is communication-intensive as it requires transmitting all sensor data to a fusion center at all times. On the other hand, the distributed \mathcal{ADLE} algorithm involves only sparse local communication among the sensors at each step, and achieves the performance of a centralized estimator asymptotically as long as the communication network stays connected in the mean. Further, note that the assumption $a = 1$ is not necessary for asymptotic normality of the \mathcal{ADLE} estimates; however, the optimality (asymptotic efficiency) is no longer guaranteed for $a \neq 1$, i.e., the resulting asymptotic covariance of the estimates deviate from Σ_c^{-1} .

4. Some Approximation Results. In this section we establish several strong (pathwise) convergence results for generic mixed time-scale stochastic recursive procedures (the proofs being provided in Appendix A). These are of independent interest and will be used in subsequent sections to analyze the properties of the \mathcal{ADLE} scheme.

Throughout this section, by $\{\mathbf{z}_t\}$, we will denote an $\{\mathcal{F}_t\}$ adapted stochastic process taking values in some Euclidean space or some subset of symmetric matrices. The initial condition \mathbf{z}_0 will be assumed to be deterministic unless otherwise stated.

Further, the probability space is assumed to be rich enough to allow the definition of various auxiliary processes governing the recursive evolution of $\{\mathbf{z}_t\}$. Since the results in this section concern generic stochastic processes not necessarily tied to the parameter vector, the θ^* indexing in the probability and expectation will be dropped temporarily.

We start by quoting a convergence rate result from [33] on deterministic recursions with time-varying coefficients.

LEMMA 4.1 (Lemmas 4 and 5 of [33]). *Let $\{\mathbf{z}_t\}$ be an \mathbb{R}_+ valued sequence satisfying*

$$\mathbf{z}_{t+1} \leq (1 - r_1(t))\mathbf{z}_t + r_2(t),$$

where $\{r_1(t)\}$ and $\{r_2(t)\}$ are deterministic sequences with

$$\frac{a_1}{(t+1)^{\delta_1}} \leq r_1(t) \leq 1 \quad \text{and} \quad r_2(t) \leq \frac{a_2}{(t+1)^{\delta_2}},$$

and $a_1 > 0$, $a_2 > 0$, $0 \leq \delta_1 \leq 1$, $\delta_2 > 0$. Then, if $\delta_1 < \delta_2$, $(t+1)^{\delta_0}\mathbf{z}_t \rightarrow 0$ as $t \rightarrow \infty$, for all $0 \leq \delta_0 < \delta_2 - \delta_1$. Also, if $\delta_1 = \delta_2$, the sequence $\{\mathbf{z}_t\}$ remains bounded, i.e., $\sup_{t \geq 0} \|\mathbf{z}_t\| < \infty$.

We now develop a stochastic analog of Lemma 4.1 in which the weight sequence $\{r_1(t)\}$ is a random process with some mixing conditions.

LEMMA 4.2. *Let $\{\mathbf{z}_t\}$ be an $\{\mathcal{F}_t\}$ adapted \mathbb{R}_+ valued process satisfying*

$$\mathbf{z}_{t+1} \leq (1 - r_1(t))\mathbf{z}_t + r_2(t).$$

In the above, $\{r_1(t)\}$ is an $\{\mathcal{F}_{t+1}\}$ adapted process, such that for all t , $r_1(t)$ satisfies $0 \leq r_1(t) \leq 1$ and

$$\frac{a_1}{(t+1)^{\delta_1}} \leq \mathbb{E}[r_1(t) \mid \mathcal{F}_t] \leq 1$$

with $a_1 > 0$ and $0 \leq \delta_1 \leq 1$. The sequence $\{r_2(t)\}$ is deterministic, \mathbb{R}_+ valued, and satisfies $r_2(t) \leq a_2/(t+1)^{\delta_2}$ with $a_2 > 0$ and $\delta_2 > 0$. Then, if $\delta_1 < \delta_2$, $(t+1)^{\delta_0}\mathbf{z}_t \rightarrow 0$ as $t \rightarrow \infty$ for all $0 \leq \delta_0 < \delta_2 - \delta_1$.

Versions of Lemma 4.2 with stronger assumptions on the weight sequences were used in earlier work. For example, the deterministic version (Lemma 4.1) was proved in [6], whereas a version with i.i.d. weight sequences was used in [33]. Further, several variants under varying assumptions exist in the literature based on generalized supermartingale convergence theorems; see for example [25, 26, 38]. However, for reasons to be clear soon, in this work there will be instances in which the memoryless assumption on the weight sequences is too restrictive. Hence, we develop the version stated in Lemma 4.2.

The following result will be used to quantify the rate of convergence of distributed vector or matrix valued recursions to their network-averaged behavior.

LEMMA 4.3. *Let $\{\mathbf{z}_t\}$ be an \mathbb{R}_+ valued $\{\mathcal{F}_t\}$ adapted process that satisfies*

$$\mathbf{z}_{t+1} \leq (1 - r_1(t))\mathbf{z}_t + r_2(t)U_t(1 + J_t).$$

Let the weight sequences $\{r_1(t)\}$ and $\{r_2(t)\}$ satisfy the hypothesis of Lemma 4.2. Further, let $\{U_t\}$ and $\{J_t\}$ be \mathbb{R}_+ valued $\{\mathcal{F}_t\}$ and $\{\mathcal{F}_{t+1}\}$ adapted processes respectively with $\sup_{t \geq 0} \|U_t\| < \infty$ a.s. The process $\{J_t\}$ is i.i.d. with J_t independent of \mathcal{F}_t for

each t and satisfies the moment condition $\mathbb{E} \left[\|J_t\|^{2+\varepsilon_1} \right] < \kappa < \infty$ for some $\varepsilon_1 > 0$ and a constant $\kappa > 0$. Then, for every δ_0 such that

$$0 \leq \delta_0 < \delta_2 - \delta_1 - \frac{1}{2 + \varepsilon_1},$$

we have $(t+1)^{\delta_0} \mathbf{z}_t \rightarrow 0$ a.s. as $t \rightarrow \infty$.

The key difference between Lemma 4.3 and Lemma 4.2 is that the processes associated with the sequence $\{r_2(t)\}$ are now stochastic.

LEMMA 4.4. Let $\{\mathbf{z}_t\}$ be an \mathbb{R}^{NM} valued $\{\mathcal{F}_t\}$ adapted process such that $\mathbf{z}_t \in \mathcal{C}^\perp$ (see (B.11) in Appendix B for the definition of the consensus subspace \mathcal{C} and its orthogonal complement \mathcal{C}^\perp) for all t . Also, let $\{L_t\}$ be an i.i.d. sequence of Laplacian matrices as in assumption (A.2) that satisfies

$$\lambda_2(\bar{L}) = \lambda_2(\mathbb{E}[L_t]) > 0,$$

with L_t being \mathcal{F}_{t+1} adapted and independent of \mathcal{F}_t for all t . Then there exists a measurable $\{\mathcal{F}_{t+1}\}$ adapted \mathbb{R}_+ valued process $\{r_t\}$ (depending on $\{\mathbf{z}_t\}$ and $\{L_t\}$) and a constant $c_r > 0$, such that $0 \leq r_t \leq 1$ a.s. and

$$\|(I_{NM} - \beta_t L_t \otimes I_M) \mathbf{z}_t\| \leq (1 - r_t) \|\mathbf{z}_t\|$$

with

$$\mathbb{E}[r_t \mid \mathcal{F}_t] \geq \frac{c_r}{(t+1)^{\tau_2}} \quad \text{a.s.} \quad (4.1)$$

for all t large enough, where the weight sequence $\{\beta_t\}$ and τ_2 are defined in (2.5).

REMARK 4.1. We comment on the necessity of the various technicalities in the statement of Lemma 4.4. Let \mathcal{P}_{NM} denote the matrix $(1/N)(\mathbf{1}_N \otimes I_M)(\mathbf{1}_N \otimes I_M)^T$ and $\mathcal{P}_{NM} \mathbf{z}_t = \mathbf{0}$ since $\mathbf{z}_t \in \mathcal{C}^\perp$. With this, a naive approach of showing the existence of such a process $\{r_t\}$ would be to use the submultiplicative inequality

$$\|(I_{NM} - \beta_t L_t \otimes I_M - \mathcal{P}_{NM}) \mathbf{z}_t\| \leq \|(I_{NM} - \beta_t L_t \otimes I_M - \mathcal{P}_{NM})\| \|\mathbf{z}_t\|.$$

Using properties of the Laplacian and the matrix \mathcal{P}_{NM} , it can be shown that for sufficiently large t

$$\|(I_{NM} - \beta_t L_t \otimes I_M - \mathcal{P}_{NM}) \mathbf{z}_t\| \leq (1 - \beta_t \lambda_2(L_t)) \|\mathbf{z}_t\|.$$

With this we may choose to define the desired sequence $\{r_t\}$ in Lemma 4.4 by

$$r_t = \beta_t \lambda_2(L_t) \quad (4.2)$$

for all t . Indeed, $\{r_t\}$ thus defined satisfies $0 \leq r_t \leq 1$ and (4.4) (at least for t large enough). Since L_t is independent of \mathcal{F}_t , we obtain

$$\mathbb{E}[\lambda_2(L_t) \mid \mathcal{F}_t] = \mathbb{E}[\lambda_2(L_t)] \leq \lambda_2(\bar{L}),$$

where the last inequality is a consequence of Jensen's inequality applied to the concave functional $\lambda_2(\cdot)$. Thus the hypothesis $\lambda_2(\bar{L}) > 0$ does not shed any light to whether $\mathbb{E}[\lambda_2(L_t)] > 0$ or not. Unfortunately, it turns out that in the gossip type of communication setting, in which none of the network instances are connected, $\lambda_2(L_t) = 0$

a.s. Hence, in such cases $\mathbb{E}[\lambda_2(L_t)]$ is actually 0. This in turn implies that the $\{r_t\}$ proposed in (4.2) violates the requirement (4.1) of Lemma 4.4. This necessitates an altogether different approach for constructing the desired sequence $\{r_t\}$. As shown in the following, such an r_t is no longer independent of \mathcal{F}_t , being a function of both L_t and \mathbf{z}_t in general.

5. Convergence and Asymptotic Properties. In this section we establish asymptotic properties of the \mathcal{ADLE} and the associated distributed learning and estimation processes. The key technical result concerning the adaptive gain learning process is presented in Lemma 5.1, whereas, the major convergence properties of the estimate processes concerning boundedness of the agent estimates, inter-agent estimate consensus and estimate consistency are obtained in Lemma 5.4, Lemma 5.7 and Lemma 5.9 respectively. The assumptions **(A.1)**-**(A.5)** are assumed to hold throughout the section.

The following result concerns the convergence of the online gain approximation processes $\{K_n(t)\}$ to their optimal counterparts $K_n = \bar{\Sigma}_c^{-1} H_n^T R_n^{-1}$.

LEMMA 5.1. *For each n the gain sequence $\{K_n(t)\}$ (given by (2.2)-(2.4)) converges to $K_n = \bar{\Sigma}_c^{-1} H_n^T R_n^{-1}$ a.s., i.e.,*

$$\mathbb{P}_{\theta^*} \left(\lim_{t \rightarrow \infty} K_n(t) = \bar{\Sigma}_c^{-1} H_n^T R_n^{-1} \right) = 1.$$

The proof is accomplished in terms of several intermediate steps that highlight the interaction between the dynamics of distributed collaboration and local adaptation. To this end, we first investigate the processes $\{G_n(t)\}$; see (2.4). The processes $\{G_n(t)\}$ may be viewed as approximations of the normalized Grammian and, as will be shown in the following, converge to $\bar{\Sigma}_c$. The following assertion concerns the consensus of the approximate Grammians to their network average and is stated as follows (see Appendix B for a proof):

LEMMA 5.2. *For each n we have*

$$\mathbb{P}_{\theta^*} \left(\lim_{t \rightarrow \infty} \|G_n(t) - G_{avg}(t)\| = 0 \right) = 1,$$

where $G_{avg}(t) = \frac{1}{N} \sum_{n=1}^N G_n(t)$ is the instantaneous network-averaged Grammian. On the basis of Lemma 5.2, to show the convergence of the approximate (normalized) Grammian sequences to $\bar{\Sigma}_c$, it suffices to show the convergence of the network-averaged sequence $\{G_{avg}(t)\}$ to the latter. This is obtained in the following lemma (see Appendix B for a proof).

LEMMA 5.3. *The following holds:*

$$\mathbb{P}_{\theta^*} \left(\lim_{t \rightarrow \infty} G_{avg}(t) = \bar{\Sigma}_c \right) = 1.$$

We now complete the proof of Lemma 5.1.

Proof. [Proof of Lemma 5.1] It follows from Lemma 5.2 and Lemma 5.3 that

$$\mathbb{P}_{\theta^*} \left(\lim_{t \rightarrow \infty} G_n(t) = \bar{\Sigma}_c \right) = 1 \tag{5.1}$$

for all $n = 1, \dots, N$. The assertion in Lemma 5.1 is immediate from (5.1) and the observation that $Q_n(t) \rightarrow R_n$ and $\gamma_t \rightarrow 0$ as $t \rightarrow \infty$. \square

The rest of the section is concerned with the convergence analysis of the estimate sequences $\{\mathbf{x}_n(t)\}$ generated by the \mathcal{ADLE} . Several results on the convergence behavior of the estimates are presented, culminating in the proofs of the main results of the paper in Section 6.

LEMMA 5.4. *The estimate sequences $\{\mathbf{x}_n(t)\}$ generated by the \mathcal{ADLE} algorithm (see (2.1)) are pathwise bounded, i.e., for each n , $\sup_{t \geq 0} \|\mathbf{x}_n(t)\| < \infty$ a.s.*

The proof of this Lemma involves a Lyapunov type argument. The decay rate estimates obtained in the next two Propositions (see Appendix B for proofs) are associated with certain time-varying spectral operators. They will be used in the construction of a suitable Lyapunov function needed in the proof of Lemma 5.4 given below.

PROPOSITION 5.5. *Let $\mathcal{K} = \text{diag}(K_1, \dots, K_N)$ by Lemma 5.1 and let $\mathcal{H} = \text{diag}(H_1, \dots, H_N)$. Then, there exist $\varepsilon_{\mathcal{K}} > 0$, a (deterministic) time $t_{\mathcal{K}}$ and a constant $c_{\mathcal{K}}$, such that,*

$$\mathbf{z}^T \left(\beta_t \bar{L} \otimes I_M + \alpha_t \tilde{\mathcal{K}} \mathcal{H} \right) \mathbf{z} \geq c_{\mathcal{K}} \alpha_t \|\mathbf{z}\|^2,$$

for all $t \geq t_{\mathcal{K}}$, $\mathbf{z} \in \mathbb{R}^{NM}$, and $\tilde{\mathcal{K}}$ satisfying $\|\tilde{\mathcal{K}} \mathcal{H} - \mathcal{K} \mathcal{H}\| \leq \varepsilon_{\mathcal{K}}$.

PROPOSITION 5.6. *Let \mathcal{K} and \mathcal{H} be defined as in Proposition 5.5. Then, for every $0 < \varepsilon < 1$ there exist a deterministic time t_{ε} and a constant c_{ε} , such that,*

$$\mathbf{z}^T \left(\beta_t \bar{L} \otimes I_M + \alpha_t \tilde{\mathcal{K}} \mathcal{H} \right) \mathbf{z} \geq c_{\varepsilon} \beta_t \|\mathbf{z}_{\mathcal{C}^{\perp}}\|^2$$

for all $t \geq t_{\varepsilon}$, $\mathbf{z} \in \mathbb{R}^{NM}$ and $\tilde{\mathcal{K}}$ satisfying

$$\left\| \tilde{\mathcal{K}} \mathcal{H} - \mathcal{K} \mathcal{H} \right\| \leq \varepsilon. \quad (5.2)$$

Also, in the above $\mathbf{z}_{\mathcal{C}^{\perp}}$ denotes the projection of \mathbf{z} in the orthogonal complement of the consensus subspace \mathcal{C} as defined in (B.11) in Appendix B.

Proof. [Proof of Lemma 5.4] The estimator recursions in (2.1) may be written as

$$\mathbf{x}_{t+1} = (I_{NM} - \beta_t \bar{L} \otimes I_M - \alpha_t \mathcal{K}_t \mathcal{H}) \mathbf{x}_t - \beta_t \left(\tilde{L}_t \otimes I_M \right) \mathbf{x}_t + \alpha_t \mathcal{K}_t \mathbf{y}_t,$$

with \mathbf{x}_t and \mathbf{y}_t denoting $[\mathbf{x}_1^T(t), \dots, \mathbf{x}_N^T(t)]^T$ and $[\mathbf{y}_1^T(t), \dots, \mathbf{y}_N^T(t)]^T$ respectively. The sequence $\{\tilde{L}_t\}$ denotes the sequence of zero mean i.i.d. matrices given by $\tilde{L}_t = L_t - \bar{L}_t$ for all t . The process $\{\mathbf{z}_t\}$ defined as $\mathbf{z}_t = \mathbf{x}_t - \mathbf{1}_N \otimes \theta^*$ may then be shown to satisfy the recursion

$$\mathbf{z}_{t+1} = (I_{NM} - \beta_t \bar{L} \otimes I_M - \alpha_t \mathcal{K}_t \mathcal{H}) \mathbf{z}_t - \beta_t \left(\tilde{L}_t \otimes I_M \right) \mathbf{z}_t + \alpha_t \mathcal{K}_t \zeta_t,$$

with $\zeta_t = [\zeta_1^T(t), \dots, \zeta_N^T(t)]^T$. Now fix $0 < \varepsilon < \varepsilon_{\mathcal{K}} \wedge 1$, where $\varepsilon_{\mathcal{K}}$ is defined in the hypothesis of Proposition 5.5. Since, $\mathcal{K}_t \rightarrow \mathcal{K}$ a.s., by Egorov's theorem ([39]) for every $\delta > 0$, there exists t_{δ} such that

$$\mathbb{P}_{\theta^*} \left(\sup_{t \geq t_{\delta}} \|\mathcal{K}_t \mathcal{H} - \mathcal{K} \mathcal{H}\| \leq \varepsilon \right) > 1 - \delta \quad \text{and} \quad \mathbb{P}_{\theta^*} \left(\sup_{t \geq t_{\delta}} \|\mathcal{K}_t - \mathcal{K}\| \leq \varepsilon \right) > 1 - \delta.$$

Moreover, such a t_{δ} may be chosen to satisfy $t_{\delta} > t_{\mathcal{K}} \vee t_{\varepsilon}$, where $t_{\mathcal{K}}$ and t_{ε} are defined in the hypotheses of Proposition 5.5 and Proposition 5.6, respectively.

Let $\mathcal{K}_{\varepsilon}$ be a (deterministic) matrix, such that,

$$\|\mathcal{K}_{\varepsilon} \mathcal{H} - \mathcal{K} \mathcal{H}\| < \varepsilon \quad \text{and} \quad \|\mathcal{K}_{\varepsilon} - \mathcal{K}\| < \varepsilon.$$

Then, for every $\delta > 0$, we may define the $\{\mathcal{F}_t\}$ adapted process $\{\mathcal{K}_t^\delta\}$, such that,

$$\mathcal{K}_t^\delta = \begin{cases} \mathcal{K}_t & \text{if } t < t_\delta \\ \mathcal{K}_t & \text{if } t \geq t_\delta \text{ and } \|\mathcal{K}_t \mathcal{H} - \mathcal{K} \mathcal{H}\| \vee \|\mathcal{K}_t - \mathcal{K}\| \leq \varepsilon \\ \mathcal{K}_\varepsilon & \text{otherwise.} \end{cases}$$

Also, for each $\delta > 0$, we define the $\{\mathcal{F}_t\}$ adapted process $\{\mathbf{z}_t^\delta\}$ by the recursion

$$\mathbf{z}_{t+1}^\delta = (I_{NM} - \beta_t \bar{L} \otimes I_M - \alpha_t \mathcal{K}_t^\delta \mathcal{H}) \mathbf{z}_t^\delta - \beta_t (\tilde{L}_t \otimes I_M) \mathbf{z}_t^\delta + \alpha_t \mathcal{K}_t^\delta \zeta_t,$$

with $\mathbf{z}_0^\delta = \mathbf{z}_0$. To show that the process $\{\mathbf{z}_t\}$ (and, hence $\{\mathbf{x}_t\}$) is bounded a.s., we note that it suffices to show that the process $\{\mathbf{z}_t^\delta\}$ is bounded a.s. for each $\delta > 0$. This is due to the fact that, by the definition of t_δ , for each $\delta > 0$ we have

$$\mathbb{P}_{\theta^*} \left(\sup_{t \geq 0} \|\mathcal{K}_t^\delta - \mathcal{K}_t\| = 0 \right) > 1 - \delta,$$

and, hence

$$\mathbb{P}_{\theta^*} \left(\sup_{t \geq 0} \|\mathbf{z}_t^\delta - \mathbf{z}_t\| = 0 \right) > 1 - \delta.$$

Thus the boundedness of the processes $\{\mathbf{z}_t^\delta\}$ for each $\delta > 0$ would imply

$$\mathbb{P}_{\theta^*} \left(\sup_{t \geq 0} \|\mathbf{x}_t\| < \infty \right) > 1 - \delta$$

for every $\delta > 0$. The assertion of Lemma 5.4 would then follow by taking δ to zero.

Hence, in the following, we focus only on the processes $\{\mathbf{z}_t^\delta\}$ and show that the latter are bounded a.s. for every $\delta > 0$. To this end, fix $\delta > 0$ and consider the \mathcal{F}_t process $V_t^\delta = \|\mathbf{z}_t^\delta\|^2$. It can be shown (Assumption **(A.3)**) that

$$\begin{aligned} \mathbb{E}_{\theta^*} [V_{t+1}^\delta | \mathcal{F}_t] &= V_t^\delta + \beta_t^2 (\mathbf{z}_t^\delta)^T \mathbb{E}_{\theta^*} \left[(\tilde{L}_t \otimes I_M)^2 \right] \mathbf{z}_t^\delta + \alpha_t^2 \mathbb{E}_{\theta^*} \left[\|\mathcal{K}_t^\delta \zeta_t\|^2 \right] \\ &\quad - 2 (\mathbf{z}_t^\delta)^T (\beta_t \bar{L} \otimes I_M + \alpha_t \mathcal{K}_t^\delta \mathcal{H}) \mathbf{z}_t^\delta + \beta_t^2 (\mathbf{z}_t^\delta)^T (\bar{L} \otimes I_M)^2 \mathbf{z}_t^\delta \\ &\quad + \alpha_t^2 (\mathbf{z}_t^\delta)^T (\mathcal{K}_t^\delta \mathcal{H})^T \mathcal{K}_t^\delta \mathcal{H} \mathbf{z}_t^\delta + 2\alpha_t \beta_t (\mathbf{z}_t^\delta)^T (\bar{L} \otimes I_M) (\mathcal{K}_t^\delta \mathcal{H}) \mathbf{z}_t^\delta. \end{aligned}$$

Since the Laplacians are bounded matrices by definition and the matrix \mathcal{K}_t^δ is bounded for $t \geq t_\delta$ by construction, there exists a constant $c_3 > 0$, sufficiently large, such that the inequalities

$$(\mathbf{z}_t^\delta)^T \mathbb{E}_{\theta^*} \left[(\tilde{L}_t \otimes I_M)^2 \right] \mathbf{z}_t^\delta = \left(\mathbf{z}_{t,C^\perp}^\delta \right)^T \mathbb{E}_{\theta^*} \left[(\tilde{L}_t \otimes I_M)^2 \right] \mathbf{z}_{t,C^\perp}^\delta \leq c_3 \left\| \mathbf{z}_{t,C^\perp}^\delta \right\|^2, \quad (5.4)$$

$$(\mathbf{z}_t^\delta)^T (\bar{L} \otimes I_M)^2 \mathbf{z}_t^\delta = \left(\mathbf{z}_{t,C^\perp}^\delta \right)^T (\bar{L} \otimes I_M)^2 \mathbf{z}_{t,C^\perp}^\delta \leq c_3 \left\| \mathbf{z}_{t,C^\perp}^\delta \right\|^2,$$

$$(\mathbf{z}_t^\delta)^T (\bar{L} \otimes I_M) (\mathcal{K}_t^\delta \mathcal{H}) \mathbf{z}_t^\delta \leq c_3 \left\| \mathbf{z}_t^\delta \right\|^2,$$

$$(\mathbf{z}_t^\delta)^T (\mathcal{K}_t^\delta \mathcal{H})^T \mathcal{K}_t^\delta \mathcal{H} \mathbf{z}_t^\delta \leq c_3 \left\| \mathbf{z}_t^\delta \right\|^2, \quad \mathbb{E}_{\theta^*} \left[\|\mathcal{K}_t^\delta \zeta_t\|^2 \right] \leq c_3 \quad (5.5)$$

hold for all $t \geq t_\delta$ with $\mathbf{z}_{t,\mathcal{C}^\perp}^\delta$ denoting the projection of \mathbf{z}_t^δ on the subspace \mathcal{C}^\perp . Also, by Proposition 5.5 and Proposition 5.6, for $t \geq t_\delta$,

$$(\mathbf{z}_t^\delta)^T (\beta_t \bar{L} \otimes I_M + \alpha_t \mathcal{K}_t^\delta \mathcal{H}) \mathbf{z}_t^\delta \geq c_{\mathcal{K}} \alpha_t \|\mathbf{z}_t^\delta\|^2 + c_\varepsilon \beta_t \left\| \mathbf{z}_{t,\mathcal{C}^\perp}^\delta \right\|^2,$$

where the positive constants $c_{\mathcal{K}}$ and c_ε are defined in the hypotheses of Proposition 5.5 and Proposition 5.6, respectively. The inequalities (5.4)-(5.5) then lead to

$$\mathbb{E}_{\theta^*} [V_{t+1}^\delta \mid \mathcal{F}_t] \leq V_t - (c_{\mathcal{K}} \beta_t - 2c_3 \beta_t^2) \left\| \mathbf{z}_{t,\mathcal{C}^\perp}^\delta \right\|^2$$

for all $t \geq t_\delta$. Observing the decay rates of the various terms in (2.5), we conclude that there exists $\bar{t}_\delta \geq t_\delta$, such that,

$$c_{\mathcal{K}} \beta_t - 2c_3 \beta_t^2 > 0 \quad \text{and} \quad c_\varepsilon \alpha_t - 2\alpha_t \beta_t c_3 - \alpha_t^2 c_3 > 0,$$

for $t \geq \bar{t}_\delta$ and, hence,

$$\mathbb{E}_{\theta^*} [V_{t+1}^\delta \mid \mathcal{F}_t] \leq V_t^\delta + c_3 \alpha_t^2 \tag{5.6}$$

for all $t \geq \bar{t}_\delta$. Let us introduce the $\{\mathcal{F}_t\}$ adapted process $\{\bar{V}_t^\delta\}$, such that,

$$\bar{V}_t^\delta = V_t^\delta - c_3 \sum_{s=t}^{\infty} \alpha_s^2 \tag{5.7}$$

for $t \geq 0$. The process $\{\bar{V}_t^\delta\}$ is well-defined as the sequence $\{\alpha_t\}$ is square summable. From (5.6) it follows immediately that

$$\mathbb{E}_{\theta^*} [\bar{V}_{t+1}^\delta \mid \mathcal{F}_t] \leq V_t^\delta - c_3 \alpha_t^2 - c_3 \sum_{s=t+1}^{\infty} \alpha_s^2 = \bar{V}_t^\delta$$

for $t \geq \bar{t}_\delta$. Hence, the process $\{\bar{V}_t^\delta\}_{t \geq \bar{t}_\delta}$ is a supermartingale. Moreover, it is bounded from below, since $V_t \geq 0$ by construction, and, in fact,

$$\bar{V}_t^\delta \geq -c_3 \sum_{s=0}^{\infty} \alpha_s^2$$

for all $t \geq 0$. Thus $\{\bar{V}_t^\delta\}_{t \geq \bar{t}_\delta}$ is a supermartingale that is bounded from below and, hence converges a.s. to a finite random variable \bar{V}^δ , i.e., $\bar{V}_t^\delta \rightarrow \bar{V}^\delta$ a.s. as $t \rightarrow \infty$. In particular, the process $\{\bar{V}_t^\delta\}$ is pathwise bounded. By (5.7) the process $\{V_t^\delta\}$ is also pathwise bounded. Thus, for each $\delta > 0$, the process $\{\mathbf{z}_t^\delta\}$ is bounded a.s. and the assertion follows. \square

The next result (see Appendix B for a proof) quantifies the rate at which the different agent estimates reach agreement and is stated as follows:

LEMMA 5.7. *For every τ_0 such that $0 \leq \tau_0 < \tau_1 - \tau_2 - 1/(2 + \varepsilon_1)$, we have*

$$\mathbb{P}_{\theta^*} \left(\lim_{t \rightarrow \infty} (t+1)^{\tau_0} (\mathbf{x}_n(t) - \mathbf{x}_{avg}(t)) = 0 \right) = 1$$

with $\mathbf{x}_{avg}(t) = (1/N) \sum_{n=1}^N \mathbf{x}_n(t)$ denoting the instantaneous network averaged estimate.

The rest of the section focuses on the convergence properties of the network averaged estimate $\{\mathbf{x}_{\text{avg}}(t)\}$ and completes the final steps required to establish the convergence properties of the agent estimates $\{\mathbf{x}_n(t)\}$. The first result in this direction concerns the consistency of the average estimate sequence.

LEMMA 5.8. *Under the additional assumption that $\tau_1 = 1$ (see (A.5)) we have*

$$\mathbb{P}_{\theta^*} \left(\lim_{t \rightarrow \infty} (\mathbf{x}_{\text{avg}}(t) - \theta^*) = 0 \right) = 1$$

with $\mathbf{x}_{\text{avg}}(t) = (1/N) \sum_{n=1}^N \mathbf{x}_n(t)$ the instantaneous network averaged estimate.

Proof. Let us denote by \mathbf{z}_t the residual $\mathbf{x}_{\text{avg}}(t) - \theta^*$. The \mathcal{F}_t -adapted process $\{\mathbf{z}_t\}$ may be shown to satisfy the recursion

$$\mathbf{z}_{t+1} = (I_M - \alpha_t \Gamma_t) \mathbf{z}_t + \alpha_t U_t + \alpha_t J_t \quad (5.8)$$

with $\{\Gamma_t\}$, $\{U_t\}$ being \mathcal{F}_t -adapted, $\{J_t\}$ being \mathcal{F}_{t+1} -adapted and given by

$$\Gamma_t = \frac{1}{N} \sum_{n=1}^N K_n(t) H_n, \quad U_t = \frac{1}{N} \sum_{n=1}^N K_n(t) (\mathbf{x}_n(t) - \mathbf{x}_{\text{avg}}(t)) \quad \text{and} \quad J_t = \frac{1}{N} K_n(t) \zeta_n(t) \quad (5.9)$$

respectively. Now fix $0 < \tau_0 < \tau_1 - \tau_2 - 1/(2 + \varepsilon_1)$ and, by the convergence of the gain processes and Lemma 5.7, $\Gamma_t \rightarrow I_M$ and $(t+1)^{\tau_0} U_t \rightarrow 0$ a.s. as $t \rightarrow \infty$. By Egorov's theorem the a.s. convergence may be assumed to be uniform on sets of arbitrarily large probability measure and, hence, for every $\delta > 0$, there exist uniformly bounded processes $\{\Gamma_t^\delta\}$, $\{U_t^\delta\}$, and $\{\mathcal{K}_t^\delta\}$ satisfying

$$\mathbb{P}_{\theta^*} \left(\sup_{s \geq t_\varepsilon^\delta} \|\Gamma_s^\delta - I_M\| \vee \|\mathcal{K}_t^\delta - \mathcal{K}\| > \varepsilon \right) = 0 \quad \text{and} \quad \mathbb{P}_{\theta^*} \left(\sup_{s \geq t_\varepsilon^\delta} (s+1)^{\tau_0} \|U_s^\delta\| > \varepsilon \right) = 0$$

for each $\varepsilon > 0$ and some t_ε^δ (sufficiently large), such that

$$\mathbb{P}_{\theta^*} \left(\sup_{t \geq 0} \|\Gamma_t^\delta - \Gamma_t\| \vee \|\mathcal{K}_t^\delta - \mathcal{K}_t\| \vee \|U_t^\delta - U_t\| = 0 \right) > 1 - \delta.$$

Also, for each $\delta > 0$, define the \mathcal{F}_t -adapted process $\{\mathbf{z}_t^\delta\}$ by

$$\mathbf{z}_{t+1}^\delta = (I_M - \alpha_t \Gamma_t^\delta) \mathbf{z}_t^\delta + \alpha_t U_t^\delta + \alpha_t J_t^\delta \quad (5.10)$$

with $\mathbf{z}_0^\delta = \mathbf{z}_0$ and $J_t^\delta = \frac{1}{N} \sum_{n=1}^N K_n^\delta(t) \zeta_n(t)$ and

$$\mathbb{P}_{\theta^*} \left(\sup_{t \geq 0} \|\mathbf{z}_t^\delta - \mathbf{z}_t\| = 0 \right) > 1 - \delta. \quad (5.11)$$

By the above development, to show that $\mathbf{z}_t \rightarrow 0$ as $t \rightarrow \infty$, it suffices to show that $\mathbf{z}_t^\delta \rightarrow 0$ as $t \rightarrow \infty$ for each $\delta > 0$. Hence, in the following, we focus on the process $\{\mathbf{z}_t^\delta\}$ only for a fixed but arbitrary $\delta > 0$.

Now let $\{V_t^\delta\}$ denote the $\{\mathcal{F}_t\}$ adapted process such that $V_t^\delta = \|\mathbf{z}_t^\delta\|^2$ for all t . Using the fact that $\mathbb{E}_{\theta^*} [\zeta_t | \mathcal{F}_t] = \mathbf{0}$ for all t , it follows that

$$\begin{aligned} \mathbb{E}_{\theta^*} [V_{t+1}^\delta | \mathcal{F}_t] &\leq \|I_M - \alpha_t \Gamma_t^\delta\|^2 V_t^\delta + 2\alpha_t (U_t^\delta)^T (I_M - \alpha_t \Gamma_t^\delta) \mathbf{z}_t^\delta \\ &\quad + \alpha_t^2 \|U_t^\delta\|^2 + \alpha_t^2 \mathbb{E}_{\theta^*} [\|J_t^\delta\|^2 | \mathcal{F}_t]. \end{aligned}$$

For t large enough

$$|2\alpha_t U_t^T (I_M - \alpha_t \Gamma_t^\delta) \mathbf{z}_t^\delta| \leq 2\alpha_t \|U_t^\delta\| \|\mathbf{z}_t^\delta\| \leq 2\alpha_t \|U_t^\delta\| \|\mathbf{z}_t^\delta\|^2 + 2\alpha_t \|U_t^\delta\|. \quad (5.13)$$

Then making t_ε^δ larger (if necessary), such that $\|U_t^\delta\| \leq \varepsilon(t+1)^{-\tau_0}$, $\mathbb{E}_{\theta^*}[\|J_t\|^2 | \mathcal{F}_t]$ is uniformly bounded, and (5.13) holds for all $t \geq t_\varepsilon^\delta$, it follows from (5.12)-(5.13) that there exist positive constants c_1 and c_2 so that

$$\begin{aligned} \mathbb{E}_{\theta^*} [V_{t+1}^\delta | \mathcal{F}_t] &\leq (1 - c_1\alpha_t + c_2\alpha_t(t+1)^{-\tau_0}) V_t^\delta \\ &\quad + c_2 (\alpha_t(t+1)^{-\tau_0} + \alpha_t^2(t+1)^{-2\tau_0} + \alpha_t^2) \end{aligned}$$

for all $t \geq t_\varepsilon^\delta$. Since $0 < \tau_0 < \tau_1$, the first term inside the second set of parenthesis on the right hand side dominates; by making $c_4 > c_2$ and $c_3 < c_1$ appropriately, we have

$$\mathbb{E}_{\theta^*} [V_{t+1}^\delta | \mathcal{F}_t] \leq (1 - c_3\alpha_t) V_t^\delta + c_4\alpha_t(t+1)^{-\tau_0} \leq V_t^\delta + c_4\alpha_t(t+1)^{-\tau_0} \quad (5.14)$$

for all $t \geq t_\varepsilon^\delta$. Now consider the $\{\mathcal{F}_t\}$ adapted process $\{\bar{V}_t^\delta\}$, such that,

$$\bar{V}_t^\delta = V_t^\delta - c_4 \sum_{s=t}^{\infty} \alpha_s (s+1)^{-\tau_0} \quad (5.15)$$

for $t \geq 0$. Since $\tau_1 = 1$ and $\tau_0 > 0$, the sequence $\{\alpha_t(t+1)^{-\tau_0}\}$ is summable and the process $\{\bar{V}_t^\delta\}$ is bounded from below. It is readily seen that $\{\bar{V}_t^\delta\}_{t \geq t_\varepsilon^\delta}$ is a supermartingale and, hence converges a.s. to a finite random variable. By (5.15), the process $\{V_t^\delta\}$ also converges a.s. to a finite random variable V^δ (necessarily non-negative). Finally, from (5.14),

$$\mathbb{E}_{\theta^*} [V_{t+1}^\delta] \leq (1 - c_3\alpha_t) \mathbb{E}_{\theta^*} [V_t^\delta] + c_4\alpha_t(t+1)^{-\tau_0}$$

for $t \geq t_\varepsilon^\delta$. Since $\tau_0 > 0$ the sequence $\{\alpha_t(t+1)^{-\tau_0}\}$ decays faster than $\{\alpha_t\}$ and, hence by Lemma 4.1 we have $\mathbb{E}_{\theta^*} [V_t^\delta] \rightarrow 0$ as $t \rightarrow \infty$. The sequence $\{V_t^\delta\}$ is non-negative, so by Fatou's lemma we further conclude that

$$0 \leq \mathbb{E}_{\theta^*} [V^\delta] \leq \liminf_{t \rightarrow \infty} \mathbb{E}_{\theta^*} [V_t^\delta] = 0.$$

The above implies $V^\delta = 0$ a.s. by the non-negativity of V^δ . Hence $\|\mathbf{z}_t^\delta\| \rightarrow 0$ as $t \rightarrow \infty$ and the desired assertion follows. \square

By inductive reasoning, we now obtain a stronger version of Lemma 5.8 that quantifies the convergence rate in the above (see Appendix B for a proof).

LEMMA 5.9. *Let assumptions (A.1)-(A.5) hold with $\tau_1 = 1$ and $a \geq 1$. Then, for each n and $\tau \in [0, 1/2)$,*

$$\mathbb{P}_{\theta^*} \left(\lim_{t \rightarrow \infty} (t+1)^\tau \|\mathbf{x}_n(t) - \theta^*\| = 0 \right) = 1. \quad (5.16)$$

6. Proofs of Main Results. The proof of Theorem 3.1 is a direct consequence of the triangle inequality and Lemma 5.7 since all agent estimates converge to the network-averaged estimate at the required rate.

Proof of Theorem 3.2

Proof. Since $\varepsilon_1 > 0$, $\tau_1 = 1$ and $\tau_1 > \tau_2 + 1/(2 + \varepsilon_1) + 1/2$, from Lemma 5.7 there exists $\varepsilon > 0$ (sufficiently small) such that

$$\mathbb{P}_{\theta^*} \left(\lim_{t \rightarrow \infty} (t+1)^{1/2+\varepsilon} \|\mathbf{x}_n(t) - \mathbf{x}_{\text{avg}}(t)\| = 0 \right) = 1$$

for all n . Moreover, by Lemma 5.9, for each $\tau \in [0, 1/2)$, we have $(t+1)^\tau \|\mathbf{x}_{\text{avg}}(t) - \theta^*\| \rightarrow 0$ a.s. as $t \rightarrow \infty$, for all n . Since $\tau < 1/2 + \varepsilon$, an immediate application of the triangle inequality yields the required estimate convergence rate. \square

Proof of Theorem 3.3

We will use the following result from [40] concerning the asymptotic normality of non-Markov stochastic recursions. The statement here is somewhat less general than in [40] but serves our application and eases the additional notational complexity.

LEMMA 6.1 (Theorem 2.2. in [40]). *Let $\{\mathbf{z}_t\}$ be an \mathbb{R}^k valued $\{\mathcal{F}_t\}$ adapted process that satisfies*

$$\mathbf{z}_{t+1} = \left(I_k - \frac{1}{t+1} \Gamma_t \right) \mathbf{z}_t + (t+1)^{-1} \Phi_t V_t + (t+1)^{-3/2} T_t,$$

where $\{V_t\}$ and $\{T_t\}$ are \mathbb{R}^k valued stochastic processes, such that, for each t , V_{t-1} and T_t are \mathcal{F}_t -adapted, and where the processes $\{\Gamma_t\}$ and $\{\Phi_t\}$ are $\mathbb{R}^{k \times k}$ valued and $\{\mathcal{F}_t\}$ adapted. Assume

$$\Gamma_t \rightarrow I_k, \quad \Phi_t \rightarrow \Phi \quad \text{and} \quad T_t \rightarrow \mathbf{0} \quad \text{as } t \rightarrow \infty.$$

Let the sequence $\{V_t\}$ satisfy $\mathbb{E}[V_t | \mathcal{F}_t] = \mathbf{0}$ for each t and there exist a constant $C > 0$ and a matrix Σ such that $C > \|\mathbb{E}[V_t V_t^T | \mathcal{F}_t] - \Sigma\| \rightarrow 0$ as $t \rightarrow \infty$, and, with

$$\sigma_{t,r}^2 = \int_{\|V_t\|^2 \geq r(t+1)} \|V_t\|^2 d\mathbb{P}, \quad (6.1)$$

let $\lim_{t \rightarrow \infty} \frac{1}{t+1} \sum_{s=0}^t \sigma_{s,r}^2 = 0$ for every $r > 0$. Then, the asymptotic distribution of $(t+1)^{-1/2} \mathbf{z}_t$ is normal with mean $\mathbf{0}$ and covariance matrix $\Phi \Sigma \Phi^T$.

Proof. [Proof of Theorem 3.3] The residual process $\{\mathbf{z}_t\}$ and its δ -approximations $\{\mathbf{z}_t^\delta\}$ are given in (5.8)-(5.10). With $\tau_1 = a = 1$,

$$\mathbf{z}_{t+1} = \left(I_M - \frac{1}{t+1} \Gamma_t \right) \mathbf{z}_t + (t+1)^{-1} U_t + (t+1)^{-1} J_t,$$

where U_t and J_t are defined in (5.8)-(5.10). Since $J_t = (1/N) \sum_{n=1}^N K_n(t) \zeta_n(t)$ and the $\{K_n(t)\}$'s may not converge uniformly (both in time and space), Lemma 6.1 is not applicable directly. Hence, we first consider the process $\{\mathbf{z}_t^\delta\}$ for some $\delta > 0$. In order to apply Lemma 6.1 to the process $\{\mathbf{z}_t^\delta\}$, define

$$T_t = (t+1)^{1/2} U_t^\delta$$

for each t . Note that by (B.22) $\|U_t^\delta\| = o((t+1)^{-1/2})$ and, hence $T_t \rightarrow \mathbf{0}$ as $t \rightarrow \infty$. Also define

$$\Phi_t = I_M \quad \text{and} \quad V_t = J_t^\delta$$

for each t . Clearly, $\mathbb{E}_{\theta^*}[V_t | \mathcal{F}_t] = \mathbf{0}$ for all t . By the convergence of \mathcal{K}_t^δ to \mathcal{K} ,

$$\lim_{t \rightarrow \infty} \mathbb{E}_{\theta^*} [V_t V_t^T | \mathcal{F}_t] = \lim_{t \rightarrow \infty} \frac{1}{N^2} \sum_{n=1}^N K_n^\delta(t) R_n (K_n^\delta)^T = \Sigma_c^{-1},$$

where the last step follows from Lemma 5.1. Moreover the uniform boundedness of the process $\{\mathcal{K}_t^\delta\}$ implies the existence of a constant $C > 0$ such that

$$\|\mathbb{E}_{\theta^*} [V_t V_t^T \mid \mathcal{F}_t] - \Sigma_c^{-1}\| < C$$

for all $t \geq 0$. The $\{V_t\}$ thus constructed also satisfies the uniform integrability assumption (6.1) due to the i.i.d. nature of the noise processes and the uniform boundedness of $\{\mathcal{K}_t^\delta\}$. Thus, the process $\{\mathbf{z}_t^\delta\}$ falls under the purview of Lemma 6.1 with $\Phi = I_M$ and $\Sigma = \Sigma_c^{-1}$. We thus conclude that

$$(t+1)^{-1/2} \mathbf{z}_t^\delta \implies \mathcal{N}(\mathbf{0}, \Sigma_c^{-1})$$

for each $\delta > 0$. To extend this asymptotic normality to the process $\{\mathbf{z}_t\}$, consider any bounded continuous function $f : \mathbb{R}^M \mapsto \mathbb{R}$. By weak convergence (Portmanteau's theorem, [41]) we have

$$\lim_{t \rightarrow \infty} \mathbb{E}_{\theta^*} \left[f \left((t+1)^{-1/2} \mathbf{z}_t^\delta \right) \right] = \mathbb{E}_{\theta^*} [f(\mathbf{z}^*)] \quad (6.2)$$

for each δ , where \mathbf{z}^* denotes a $\mathcal{N}(\mathbf{0}, \Sigma_c^{-1})$ distributed random vector under the measure \mathbb{P}_* . Denoting by $\|f\|_\infty$ the sup-norm of $f(\cdot)$ (necessarily finite) we obtain from (5.11)

$$\left\| \mathbb{E}_{\theta^*} \left[f \left((t+1)^{-1/2} \mathbf{z}_t^\delta \right) \right] - \mathbb{E}_{\theta^*} \left[f \left((t+1)^{-1/2} \mathbf{z}_t \right) \right] \right\| \leq 2\delta \|f\|_\infty.$$

By (6.2) we then have

$$\limsup_{t \rightarrow \infty} \left\| \mathbb{E}_{\theta^*} \left[f \left((t+1)^{-1/2} \mathbf{z}_t \right) \right] - \mathbb{E}_{\theta^*} [f(\mathbf{z}^*)] \right\| \leq 2\delta \|f\|_\infty.$$

Since the above holds for each $\delta > 0$, we conclude that $\mathbb{E}_{\theta^*} [f((t+1)^{-1/2} \mathbf{z}_t)] \rightarrow \mathbb{E}_{\theta^*} [f(\mathbf{z}^*)]$ as $t \rightarrow \infty$. This convergence holds for all bounded continuous functions $f(\cdot)$ thus giving the required weak convergence of the sequence $\{(t+1)^{-1/2} \mathbf{z}_t\}$. \square

7. Conclusion. We have developed a distributed estimator that combines a recursive collaborative learning step with the estimate update task. Through this learning process, the agents adaptively improve their quantitative model information and innovation gains with a view toward achieving the performance of the optimal centralized estimator. Intuitively, the distributed approach is a culmination of two potentials, the agreement (or consensus) and the innovation. By properly designing the relative strength of their excitations, we have shown that the agent estimates may be made asymptotically efficient in terms of their asymptotic covariance which coincides with the asymptotic covariance (the inverse of the Fisher information rate for Gaussian systems) of a centralized estimator with perfect statistical information and having access to all agent observations at all times. A typical application scenario involves multi-sensor distributed platforms, for example, the smart grid or vehicular networks. Such networks are generally equipped with rich sensing infrastructures and high sensing diversity, but suffer from lack of information about the global model and about the relative observation efficiencies due to unpredictable changes and constraints in the sensing resources. Extensions of this work to nonlinear sensing platforms are currently being investigated. Another important direction will be the extension of this adaptive collaborative scheme to dynamic parameter situations as opposed to the static case

considered in this paper.

Appendix A. Proofs in Section 4.

Proof. Proof of Lemma 4.2 We start by showing that for each positive integer k , the following holds:

$$\lim_{t \rightarrow \infty} (t+1)^{k(\delta_2 - \delta_1 - \varepsilon_0)} \mathbb{E} \left[\mathbf{z}_t^k \right] = 0 \quad (\text{A.1})$$

for every $0 < \varepsilon_0 \leq \delta_2 - \delta_1$. The proof proceeds by induction on k . Let us first consider $k = 1$. We then have

$$\begin{aligned} \mathbb{E} [\mathbf{z}_{t+1}] &\leq \mathbb{E} [(1 - \mathbb{E}[r_1(t) \mid \mathcal{F}_t]) \mathbf{z}_t] + r_2(t) \\ &\leq (1 - \bar{r}_1(t)) \mathbb{E}[\mathbf{z}_t] + r_2(t), \end{aligned}$$

where by $\bar{r}_1(t)$ we denote the quantity $a_1/(t+1)^{\delta_1}$. The deterministic \mathbb{R}_+ valued sequence $\{\mathbb{E}[\mathbf{z}_t]\}$ satisfies the conditions of Lemma 4.1 and the claim in (A.1) holds for $k = 1$. Now assume the claim in (A.1) holds for all $k \leq k_0$, with k_0 a positive integer. We now show that the claim also holds for $k = k_0 + 1$. Indeed, by the polynomial expansion

$$\mathbf{z}_{t+1}^{k_0+1} = \sum_{i=0}^{k_0+1} \binom{k_0+1}{i} ((1 - r_1(t)) \mathbf{z}_t)^{k_0+1-i} r_2^i(t)$$

and the fact that $0 \leq r_1(t) \leq 1$, we have

$$\mathbf{z}_{t+1}^{k_0+1} \leq (1 - r_1(t)) \mathbf{z}_t^{k_0+1} + \sum_{i=1}^{k_0+1} \binom{k_0+1}{i} \mathbf{z}_t^{k_0+1-i} r_2^i(t).$$

In a way similar to (A.2), the above implies

$$\mathbb{E} \left[\mathbf{z}_{t+1}^{k_0+1} \right] \leq (1 - \bar{r}_1(t)) \mathbb{E} \left[\mathbf{z}_t^{k_0+1} \right] + \sum_{i=1}^{k_0+1} \binom{k_0+1}{i} \mathbb{E} \left[\mathbf{z}_t^{k_0+1-i} \right] r_2^i(t). \quad (\text{A.3})$$

By the induction hypothesis and the assumptions on the sequence $\{r_2(t)\}$, there exist constants c_i for $i = 1, \dots, k_0 + 1$, such that,

$$\mathbb{E} \left[\mathbf{z}_t^{k_0+1-i} \right] r_2^i(t) \leq \frac{c_i}{(t+1)^{(k_0+1-i)(\delta_2 - \delta_1 - \varepsilon_0) + i\delta_2}} = \frac{c_i}{(t+1)^{(k_0+1)(\delta_2 - \delta_1 - \varepsilon_0) + i(\delta_1 + \varepsilon_0)}} \quad (\text{A.4})$$

for all $i = 1, \dots, k_0 + 1$. It is readily seen that the smallest decay rate in the above is attained at $i = 1$. Hence, from (A.3)-(A.4), there exists another constant c_0 , such that,

$$\mathbb{E} \left[\mathbf{z}_{t+1}^{k_0+1} \right] \leq (1 - \bar{r}_1(t)) \mathbb{E} \left[\mathbf{z}_t^{k_0+1} \right] + \frac{c_0}{(t+1)^{(k_0+1)(\delta_2 - \delta_1 - \varepsilon_0) + (\delta_1 + \varepsilon_0)}}.$$

The deterministic sequence $\left\{ \mathbb{E} \left[\mathbf{z}_t^{k_0+1} \right] \right\}$ then falls under the purview of Lemma 4.1 (by taking $\delta_2 \doteq (k_0 + 1)(\delta_2 - \delta_1 - \varepsilon_0) + (\delta_1 + \varepsilon_0)$ and $\delta_1 \doteq \delta_1$). Since $\varepsilon_0 > 0$, an immediate application of Lemma 4.1 gives

$$\lim_{t \rightarrow \infty} (t+1)^{(k_0+1)(\delta_2 - \delta_1 - \varepsilon_0)} \mathbb{E} \left[\mathbf{z}_t^{k_0+1} \right] = 0$$

and the induction step follows. This establishes the desired claim in (A.1).

We now complete the proof of Lemma 4.2. To this end, choose $\bar{\delta}$, such that $0 < \bar{\delta} < \delta_2 - \delta_1 - \delta_0$. Let k_{δ_0} be a positive integer, such that $k_{\delta_0}(\delta_2 - \delta_1 - \delta_0 - \bar{\delta}) > 1$. Then, for every $\varepsilon > 0$, we have

$$\mathbb{P} \left((t+1)^{\delta_0} \mathbf{z}_t > \varepsilon \right) \leq \frac{\mathbb{E}[\mathbf{z}_t^{k_{\delta_0}}]}{\varepsilon^{k_{\delta_0}} (t+1)^{-k_{\delta_0} \delta_0}} \leq \frac{c}{\varepsilon^{k_{\delta_0}} (t+1)^{k_{\delta_0}(\delta_2 - \delta_1 - \delta_0 - \bar{\delta})}}. \quad (\text{A.5})$$

The last step is a consequence of the claim in (A.1), by which there exists a constant $c > 0$, such that,

$$\mathbb{E}[\mathbf{z}_t^{k_{\delta_0}}] \leq \frac{c}{(t+1)^{k_{\delta_0}(\delta_2 - \delta_1 - \bar{\delta})}}$$

for all $t \geq 0$. Since $k_{\delta_0}(\delta_2 - \delta_1 - \delta_0 - \bar{\delta}) > 1$ by choice, the rightmost term in (A.5) is summable in t . We thus obtain $\sum_{t=0}^{\infty} \mathbb{P}((t+1)^{\delta_0} \mathbf{z}_t > \varepsilon) < \infty$, and, hence,

$$\mathbb{P}\left((t+1)^{\delta_0} \mathbf{z}_t > \varepsilon \text{ i.o.}\right) = 0 \quad (\text{A.6})$$

by the Borel-Cantelli lemma (i.o. stands for infinitely often in (A.6)). Since (A.6) holds for arbitrary $\varepsilon > 0$, we conclude that $(t+1)^{\delta_0} \mathbf{z}_t \rightarrow 0$ a.s. as $t \rightarrow \infty$. \square

Proof. [Proof of Lemma 4.3] Fix $\delta \in \left(0, \delta_2 - \delta_1 - \delta_0 - \frac{1}{2+\varepsilon_1}\right)$. The following is readily verified:

For every $\varepsilon_3 > 0$, there exists $R_{\varepsilon_3} > 0$, such that

$$\mathbb{P}\left(\sup_{t \geq 0} \frac{1}{(t+1)^{\frac{1}{2+\varepsilon_1} + \delta}} \|U_t(1+J_t)\| < R_{\varepsilon_3}\right) > 1 - \varepsilon_3. \quad (\text{A.7})$$

Indeed, for any $\varepsilon_2 > 0$, we note that

$$\begin{aligned} \mathbb{P}\left(\frac{1}{(t+1)^{\frac{1}{2+\varepsilon_1} + \delta}} \|J_t\| > \varepsilon_2\right) &\leq \frac{1}{\varepsilon_2^{2+\varepsilon_1} (t+1)^{1+\delta(2+\varepsilon_1)}} \mathbb{E}[\|J_t\|^{2+\varepsilon_1}] \\ &\leq \frac{\kappa}{\varepsilon_2^{2+\varepsilon_1} (t+1)^{1+\delta(2+\varepsilon_1)}}. \end{aligned}$$

Since $\delta > 0$, the term on the right hand side of (A.8) is summable, and by the Borel-Cantelli lemma we may conclude that

$$\mathbb{P}\left(\frac{1}{(t+1)^{\frac{1}{2+\varepsilon_1} + \delta}} \|J_t\| > \varepsilon_2 \text{ i.o.}\right) = 0.$$

Since ε_2 is arbitrary, it follows that

$$\mathbb{P}\left(\lim_{t \rightarrow \infty} \frac{1}{(t+1)^{\frac{1}{2+\varepsilon_1} + \delta}} \|J_t\| = 0\right) = 1. \quad (\text{A.9})$$

From the boundedness of $\{U_t\}$ and (A.9) we may further conclude that

$$\mathbb{P}\left(\lim_{t \rightarrow \infty} \frac{1}{(t+1)^{\frac{1}{2+\varepsilon_1} + \delta}} \|U_t(1+J_t)\| = 0\right) = 1. \quad (\text{A.10})$$

By Egorov's theorem the a.s. convergence in (A.10) is uniform except on a set of arbitrarily small measure, which verifies the claim in (A.7).

We now establish the desired result by a truncation argument. For a scalar a , define its truncation $(a)_C$ at level $C > 0$ by

$$(a)_C = \begin{cases} \frac{a}{|a|} \min(|a|, C) & \text{if } a \neq 0 \\ 0 & \text{if } a = 0. \end{cases} \quad (\text{A.11})$$

For a vector, the truncation operation applies component-wise. Now, for each $C > 0$, consider the sequence $\{\widehat{\mathbf{z}}_C(t)\}$ given by the recursion

$$\widehat{\mathbf{z}}_C(t+1) = (1 - r_1(t))\widehat{\mathbf{z}}_C(t) + r_2(t) (U_t(1+J_t))_{C(t+1)^{\frac{1}{2+\varepsilon_1} + \delta}} \quad (\text{A.12})$$

with $\widehat{\mathbf{z}}_C(0) = \mathbf{z}_0$. Using (A.11), we have

$$\widehat{\mathbf{z}}_C(t+1) \leq (1 - r_1(t))\widehat{\mathbf{z}}_C(t) + \widehat{r}_2(t), \quad (\text{A.13})$$

where

$$\widehat{r}_2(t) \leq \frac{k_1}{(t+1)^{\delta_2 - \delta - \frac{1}{2+\varepsilon_1}}}, \quad \forall t \quad (\text{A.14})$$

for some constant $k_1 > 0$. By construction the process $\{\widehat{\mathbf{z}}_C(t)\}$ is $\{\mathcal{F}_t\}$ adapted and, hence, the recursion in (A.13)-(A.14) falls under the purview of Lemma 4.2. Thus, for every $C > 0$, we have $(t+1)^{\delta_0}\widehat{\mathbf{z}}_C(t) \rightarrow 0$ a.s. as $t \rightarrow \infty$, since $\delta_0 < \delta_2 - \delta_1 - \delta - \frac{1}{2+\varepsilon_1}$.

Now, for $\varepsilon_3 > 0$, consider the sequence $\{\widehat{\mathbf{z}}_{R_{\varepsilon_3}}(t)\}$, where $R_{\varepsilon_3} > 0$ is the constant in (A.7). Using (A.7) and (A.12) we may conclude that

$$\mathbb{P}\left(\inf_{t \geq 0} (\widehat{\mathbf{z}}_{R_{\varepsilon_3}}(t) - \mathbf{z}_t) \geq 0\right) > 1 - \varepsilon_3. \quad (\text{A.15})$$

Since all processes involved are non-negative, it readily follows from (A.15) that

$$\mathbb{P}\left(\lim_{t \rightarrow \infty} (t+1)^{\delta_0} \mathbf{z}_t = 0\right) > 1 - \varepsilon_3. \quad (\text{A.16})$$

The lemma follows by taking ε_3 to zero in (A.16). \square

Proof. [Proof of Lemma 4.4] Let \mathcal{L} denote the set of possible Laplacian matrices (necessarily finite) and \mathcal{D} the distribution on \mathcal{L} induced by the link formation process. Since the set of Laplacian matrices is finite, the set \mathcal{L} may be chosen such that $\underline{p} = \inf_{L \in \mathcal{L}} p_L > 0$, with $p_L = \mathbb{P}(L_t = L)$ for each $L \in \mathcal{L}$ and $\sum_{L \in \mathcal{L}} p_L = 1$. The hypothesis $\lambda_2(\bar{L}) > 0$ implies that for every $\mathbf{z} \in \mathcal{C}^\perp$,

$$\sum_{L \in \mathcal{L}} \mathbf{z}^T L \mathbf{z} \geq \sum_{L \in \mathcal{L}} \mathbf{z}^T (p_L L) \mathbf{z} = \mathbf{z}^T \bar{L} \mathbf{z} \geq \lambda_2(\bar{L}) \|\mathbf{z}\|^2. \quad (\text{A.17})$$

Denoting by $|\mathcal{L}|$ the cardinality of \mathcal{L} , it follows from (A.17) that for each $\mathbf{z} \in \mathcal{C}^\perp$ there exists some $L_{\mathbf{z}} \in \mathcal{L}$, such that $\mathbf{z}^T L_{\mathbf{z}} \mathbf{z} \geq (\lambda_2(\bar{L})/|\mathcal{L}|) \|\mathbf{z}\|^2$. Moreover, since the set \mathcal{L} is finite, the mapping $L_{\mathbf{z}} : \mathcal{C}^\perp \rightarrow \mathcal{L}$ may be realized as a measurable function.

For each $L \in \mathcal{L}$, the eigenvalues of the matrix $I_{NM} - \beta_t L \otimes I_M$ are 1 and $1 - \beta_t \lambda_n(L)$, $2 \leq n \leq N$, each being repeated M times. Hence, for $t \geq t_0$ (large enough), $\|I_{NM} - \beta_t L \otimes I_M\| \leq 1$ and $\|(I_{NM} - \beta_t L \otimes I_M)\mathbf{z}\| \leq \|\mathbf{z}\|$ for every $\mathbf{z} \in \mathbb{R}^{NM}$. Hence, the functional $r_{L,\mathbf{z}}$ given by

$$r_{L,\mathbf{z}} = \begin{cases} 1 & \text{if } t < t_0 \text{ or } \mathbf{z} = \mathbf{0} \\ 1 - \frac{\|(I_{NM} - \beta_t L \otimes I_M)\mathbf{z}\|}{\|\mathbf{z}\|} & \text{otherwise} \end{cases}$$

is jointly measurable in L and \mathbf{z} and satisfies $0 \leq r_{L,\mathbf{z}} \leq 1$ for each pair (L, \mathbf{z}) . Let $\{r_t\}$ be the $\{\mathcal{F}_{t+1}\}$ adapted process given by $r_t = r_{L_t, \mathbf{z}_t}$ for each t , and $\|(I_{NM} - \beta_t L \otimes I_M)\mathbf{z}_t\| = (1 - r_t)\|\mathbf{z}_t\|$ a.s. for each t . We now need to verify that $\{r_t\}$ satisfies (4.1) for some $c_r > 0$. To this end, for t large enough,

$$\begin{aligned} \|(I_{NM} - \beta_t L_{\mathbf{z}_t} \otimes I_M)\mathbf{z}_t\|^2 &= \mathbf{z}_t^T (I_{NM} - 2\beta_t L_{\mathbf{z}_t} \otimes I_M) \mathbf{z}_t + \beta_t^2 \mathbf{z}_t^T (L_{\mathbf{z}_t} \otimes I_M)^2 \mathbf{z}_t \\ &\leq (1 - 2\beta_t \lambda_2(\bar{L})/|\mathcal{L}|) \|\mathbf{z}_t\|^2 + c_1 \beta_t^2 \|\mathbf{z}_t\|^2 \\ &\leq (1 - \beta_t \lambda_2(\bar{L})/|\mathcal{L}|) \|\mathbf{z}_t\|^2, \end{aligned}$$

where we have used the definition of the function $L_{\mathbf{z}}$, the boundedness of the Laplacian matrix and the fact that $\beta_t \rightarrow 0$. Hence, by making t_0 larger if necessary, we have

$$\|(I_{NM} - \beta_t L_{\mathbf{z}_t} \otimes I_M)\mathbf{z}_t\| \leq \left(1 - \beta_t \frac{\lambda_2(\bar{L})}{4|\mathcal{L}|}\right) \|\mathbf{z}_t\| \quad (\text{A.19})$$

for all $t \geq t_0$. Now, by (A.19)

$$\begin{aligned} \mathbb{E}[\|(I_{NM} - \beta_t L \otimes I_M)\mathbf{z}_t\| \mid \mathcal{F}_t] &= \sum_{L \in \mathcal{L}} p_L (1 - r_{L, \mathbf{z}_t}) \|\mathbf{z}_t\| \\ &\leq \left(1 - \left(\underline{p}\beta_t \frac{\lambda_2(\bar{L})}{4|\mathcal{L}|} + \sum_{L \neq L_{\mathbf{z}_t}} p_L r_{L, \mathbf{z}_t}\right)\right) \|\mathbf{z}_t\|. \end{aligned}$$

Since $\sum_{L \neq L_{\mathbf{z}_t}} p_L r_{L, \mathbf{z}_t} \geq 0$, we have for $t \geq t_0$,

$$(1 - \mathbb{E}[r_t \mid \mathcal{F}_t])\|\mathbf{z}_t\| = \mathbb{E}[\|(I_{NM} - \beta_t L \otimes I_M)\mathbf{z}_t\| \mid \mathcal{F}_t] \leq \left(1 - \underline{p}\beta_t \frac{\lambda_2(\bar{L})}{4|\mathcal{L}|}\right) \|\mathbf{z}_t\|.$$

Since, by definition $r_t = 1$ on the set $\{\mathbf{z}_t = \mathbf{0}\}$, it follows that

$$\mathbb{E}[r_t \mid \mathcal{F}_t] \geq \frac{\underline{p}\lambda_2(\bar{L})}{4|\mathcal{L}|} \beta_t$$

for all $t \geq t_0$, thus establishing the assertion. \square

Appendix B. Proofs in Section 5.

Proof. [Proof of Lemma 5.2] We will show the desired convergence in the matrix Frobenius norm (denoted by $\|\cdot\|_F$ in the following). Since the matrix space under consideration is finite dimensional, the convergence in \mathcal{L}_2 norm will follow. The existence of quadratic moments implies the convergence of the sample covariances (see (2.3)) to the true covariances and, hence, for each n , $Q_n(t) \rightarrow R_n$ a.s. Since, in addition, the sequence $\{\gamma_t\}$ in (2.2) goes to zero, we may choose an a.s. finite random variable R_2 , such that for each n ,

$$\mathbb{P}_{\theta^*} \left(\sup_{t \geq 0} \left\| H_n^T (Q_n(t) + \gamma_t I_{M_n})^{-1} H_n \right\| \leq R_2 < \infty \right) = 1. \quad (\text{B.1})$$

By construction, the matrix sequences $\{G_n(t)\}$ and $\{Q_n(t)\}$ are symmetric for each n . Let $\tilde{G}_n(t) = G_n(t) - G_{\text{avg}}(t)$ denote the deviation of the Grammian estimate at agent n from the instantaneous network average $G_{\text{avg}}(t)$. Also, let \tilde{G}_t and D_t respectively denote the matrices $[\tilde{G}_1(t), \dots, \tilde{G}_N(t)]^T$ and $[D_1(t), \dots, D_N(t)]^T$, where $D_n(t) = (Q_n(t) + \gamma_t I_{M_n})^{-1}$ for each n . Using the following readily verifiable properties of the Laplacian:

$$(\mathbf{1}_N \otimes I_M)^T (L_t \otimes I_M) = 0, \quad (L_t \otimes I_M) (\mathbf{1}_N \otimes G_{\text{avg}}(t)) = \mathbf{0}, \quad (\text{B.2})$$

we have

$$\tilde{G}_{t+1} = (I_{NM} - \beta_t (L_t \otimes I_M) - \alpha_t I_{NM}) \tilde{G}_t + \alpha_t ((D_t - D_{\text{avg}}(t))), \quad (\text{B.3})$$

where $D_{\text{avg}}(t) = \frac{1}{N} \sum_{n=1}^N D_n(t)$. Note that, by (B.1), there exists an $\{\mathcal{F}_t\}$ adapted a.s. bounded process $\{U_t\}$, such that $\sup_{t \geq 0} \|D_t - D_{\text{avg}}(t)\|_F \leq U_t$ a.s. For $m \in \{1, \dots, M\}$, let $\tilde{G}_{m,t}$ denote the m -th column of \tilde{G}_t . The process $\{\tilde{G}_{m,t}\}$ is $\{\mathcal{F}_t\}$ adapted and $\tilde{G}_{m,t} \in \mathcal{C}^1$ for each t . Then, by Lemma 4.4 there exists a $[0, 1]$ -valued $\{\mathcal{F}_{t+1}\}$ adapted process $\{r_{m,t}\}$, such that,

$$\|(I_{NM} - \beta_t L_t \otimes I_M) \tilde{G}_{m,t}\| \leq (1 - r_{m,t}) \|\tilde{G}_{m,t}\|$$

and $\mathbb{E}_{\theta^*}[r_{m,t} \mid \mathcal{F}_t] \geq c_{m,r}/(t+1)^{\tau_2}$ a.s. for $t \geq t_0$ sufficiently large. Noting that the square of the Frobenius norm is the sum of the squared column \mathcal{L}_2 norms, we have

$$\|(I_{NM} - \beta_t L_t \otimes I_M) \tilde{G}_t\|_F^2 \leq \sum_{m=1}^M (1 - r_{m,t})^2 \|\tilde{G}_{m,t}\|^2 \leq (1 - r_t)^2 \|\tilde{G}_t\|_F^2, \quad (\text{B.4})$$

where $\{r_t\}$ is the $\{\mathcal{F}_{t+1}\}$ adapted process given by $r_t = r_{1,t} \wedge r_{2,t} \wedge \dots \wedge r_{M,t}$. By the conditional Jensen's inequality, we obtain

$$\mathbb{E}_{\theta^*}[r_t \mid \mathcal{F}_t] \geq \wedge_{m=1}^M \mathbb{E}_{\theta^*}[r_{m,t} \mid \mathcal{F}_t] \geq c_r/(t+1)^{\tau_2} \quad (\text{B.5})$$

for some $c_r > 0$ and $t \geq t_0$. Recall $\{\alpha_t\}$ from (2.5). Using (B.4), we finally get

$$\begin{aligned} \|(I_{NM} - \beta_t L_t \otimes I_M - \alpha_t I_{NM})\tilde{G}_t\|_F &\leq \|(I_{NM} - \beta_t L_t \otimes I_M)\tilde{G}_t\|_F + \alpha_t \|\tilde{G}_t\|_F \\ &\leq (1 - r_t) \|\tilde{G}_t\|_F + \alpha_t \|\tilde{G}_t\|_F \\ &\leq (1 - r_t/2) \|\tilde{G}_t\|_F \end{aligned} \quad (\text{B.6})$$

for $t \geq t_0$. From (B.3) and (B.6) we then have

$$\|\tilde{G}_{t+1}\|_F \leq \|(I_{NM} - \beta_t L_t \otimes I_M - \alpha_t I_{NM})\tilde{G}_t\|_F + \alpha_t U_t \leq (1 - r_t/2) \|\tilde{G}_t\|_F + \alpha_t U_t. \quad (\text{B.7})$$

By (B.5) and since $\beta_t/\alpha_t \rightarrow \infty$ as $t \rightarrow \infty$, the recursion in (B.7) clearly falls under the purview of Lemma 4.3, and we conclude that $\|\tilde{G}_t\|_F \rightarrow 0$ a.s. as $t \rightarrow \infty$. The convergence in the \mathcal{L}_2 norm follows immediately. \square

Proof. [Proof of Lemma 5.3] The process $\{G_{\text{avg}}(t)\}$ satisfies the following recursion:

$$G_{\text{avg}}(t+1) = (1 - \alpha_t)G_{\text{avg}}(t) + \alpha_t D_{\text{avg}}(t).$$

Let $\tilde{G}_{\text{avg}}(t)$ denote the residual $G_{\text{avg}}(t) - \bar{\Sigma}_c$ and the process $\{\tilde{G}_{\text{avg}}(t)\}$ satisfies

$$\tilde{G}_{\text{avg}}(t+1) = (1 - \alpha_t)\tilde{G}_{\text{avg}}(t) + \alpha_t (D_{\text{avg}}(t) - \bar{\Sigma}_c). \quad (\text{B.8})$$

By Eqn. 139, Lemma 25 in [6] there exist t_0 sufficiently large and a constant B such that

$$0 \leq \sum_{k=s}^{t-1} \left(\prod_{l=k+1}^{t-1} (1 - \alpha_l) \right) \alpha_k \leq B,$$

for all positive integers t and s with $t_0 \leq s \leq t$. Also, the convergence of the sample covariances and the fact that $\gamma_t \rightarrow 0$ as $t \rightarrow \infty$ imply $D_{\text{avg}}(T) \rightarrow \bar{\Sigma}_c$ a.s. as $t \rightarrow \infty$. Hence, for a given $\varepsilon > 0$, we may choose $t_\varepsilon > t_0$ such that $\|D_{\text{avg}}(t) - \bar{\Sigma}_c\| < \varepsilon$ for all $t \geq t_\varepsilon$. From (B.8), we then have for $t > t_\varepsilon$

$$\begin{aligned} \|\tilde{G}_{\text{avg}}(t)\| &\leq \left| \prod_{k=t_\varepsilon}^{t-1} (1 - \alpha_k) \right| \|\tilde{G}_{\text{avg}}(t_\varepsilon)\| + \sum_{k=t_\varepsilon}^{t-1} \left(\prod_{l=k+1}^{t-1} (1 - \alpha_l) \right) \alpha_k \varepsilon \\ &\leq \left| \prod_{k=t_\varepsilon}^{t-1} (1 - \alpha_k) \right| \|\tilde{G}_{\text{avg}}(t_\varepsilon)\| + B\varepsilon. \end{aligned} \quad (\text{B.9})$$

Since $\sum_{t \geq 0} \alpha_t = \infty$ the first term on the right hand side of (B.9) goes to zero as $t \rightarrow \infty$, and we have $\limsup_{t \rightarrow \infty} \|\tilde{G}_{\text{avg}}(t)\| \leq B\varepsilon$. Since $\varepsilon > 0$ is arbitrary, we conclude that $\tilde{G}_{\text{avg}}(t) \rightarrow \mathbf{0}$ a.s. as $t \rightarrow \infty$ by taking ε to zero. The desired assertion follows immediately. \square

Proof. [Proof of Proposition 5.5] A version of this result was established in [33] (Lemma 6) for the case of constant gains $K_n(t)$. In the following we generalize the arguments of [33] to time-varying adaptive gains. To this end we show

$$\inf_{\|\mathbf{z}\|=1} \mathbf{z}^T \left(\frac{\beta_t}{\alpha_t} \bar{L} \otimes I_M + \mathcal{KH} \right) \mathbf{z} > 0 \quad (\text{B.10})$$

for all t sufficiently large, where $\mathcal{K} = \text{diag}(K_1, \dots, K_N)$.

A vector $\mathbf{z} \in \mathbb{R}^{NM}$ may be decomposed as $\mathbf{z} = \mathbf{z}_c + \mathbf{z}_{c^\perp}$, with \mathbf{z}_c denoting its projection on the consensus or agreement subspace \mathcal{C} ,

$$\mathcal{C} = \left\{ \mathbf{z} \in \mathbb{R}^{NM} \mid \mathbf{z} = \mathbf{1}_N \otimes \mathbf{a} \text{ for some } \mathbf{a} \in \mathbb{R}^M \right\}, \quad (\text{B.11})$$

and \mathbf{z}_{c^\perp} the orthogonal complement. Also, denoting by $\mathcal{D}_\mathcal{K}$ the symmetricized version of \mathcal{KH} , i.e., $\mathcal{D}_\mathcal{K} = \frac{1}{2} (\mathcal{KH} + \mathcal{H}^T \mathcal{K}^T)$, standard matrix manipulations and properties of the Laplacian yield

$$\mathbf{z}^T \left(\frac{\beta_t}{\alpha_t} \bar{L} \otimes I_M + \mathcal{KH} \right) \mathbf{z} \geq \frac{\beta_t}{\alpha_t} \lambda_2(\bar{L}) \|\mathbf{z}_{c^\perp}\|^2 + \mathbf{z}_{c^\perp}^T \mathcal{D}_\mathcal{K} \mathbf{z}_{c^\perp} + 2\mathbf{z}_c^T \mathcal{D}_\mathcal{K} \mathbf{z}_{c^\perp} + \mathbf{z}_c^T \mathcal{D}_\mathcal{K} \mathbf{z}_c. \quad (\text{B.12})$$

By construction, $\sum_{n=1}^N K_n H_n = \bar{\Sigma}_c^{-1} \sum_{n=1}^N H_n^T R_n^{-1} H_n = N I_M$, and, hence, we note that $\mathbf{z}_c^T \mathcal{D}_K \mathbf{z}_c = \|\mathbf{z}_c\|^2$ for each $\mathbf{z} \in \mathbb{R}^{NM}$. Let us choose a constant $c_1 > 0$ such that

$$\mathbf{z}_{c^\perp}^T \mathcal{D}_K \mathbf{z}_{c^\perp} \geq -c_1 \|\mathbf{z}_{c^\perp}\|^2 \quad \text{and} \quad \mathbf{z}_c^T \mathcal{D}_K \mathbf{z}_{c^\perp} \geq -c_1 \|\mathbf{z}_c\| \|\mathbf{z}_{c^\perp}\|.$$

It then follows from (B.12) that

$$\mathbf{z}^T \left(\frac{\beta_t}{\alpha_t} \bar{L} \otimes I_M + \mathcal{K}\mathcal{H} \right) \mathbf{z} \geq \left(\frac{\beta_t}{\alpha_t} \lambda_2(\bar{L}) - c_1 \right) \|\mathbf{z}_{c^\perp}\|^2 - 2c_1 \|\mathbf{z}_c\| \|\mathbf{z}_{c^\perp}\| + \|\mathbf{z}_c\|^2. \quad (\text{B.13})$$

Since $\beta_t/\alpha_t \rightarrow \infty$ and $\lambda_2(\bar{L}) > 0$, there exists t_1 sufficiently large such that

$$\frac{\beta_t}{\alpha_t} \lambda_2(\bar{L}) - c_1 > c_1^2, \quad \forall t \geq t_1. \quad (\text{B.14})$$

We now verify (B.10) for $t \geq t_1$. To this end, assume $\|\mathbf{z}\| = 1$. In case $\mathbf{z}_c = \mathbf{0}$ ($\|\mathbf{z}_{c^\perp}\| = 1$), we have from (B.13)

$$\mathbf{z}^T \left(\frac{\beta_t}{\alpha_t} \bar{L} \otimes I_M + \mathcal{K}\mathcal{H} \right) \mathbf{z} \geq \frac{\beta_t}{\alpha_t} \lambda_2(\bar{L}) - c_1 > 0.$$

For the other case, i.e., $\mathbf{z}_c \neq \mathbf{0}$,

$$\mathbf{z}^T \left(\frac{\beta_t}{\alpha_t} \bar{L} \otimes I_M + \mathcal{K}\mathcal{H} \right) \mathbf{z} \geq \|\mathbf{z}_c\|^2 \left[\left(\frac{\beta_t}{\alpha_t} \lambda_2(\bar{L}) - c_1 \right) \frac{\|\mathbf{z}_{c^\perp}\|^2}{\|\mathbf{z}_c\|^2} - 2c_1 \frac{\|\mathbf{z}_{c^\perp}\|}{\|\mathbf{z}_c\|} + 1 \right] > 0,$$

where the last inequality follows from the fact that the quadratic functional of $\frac{\|\mathbf{z}_{c^\perp}\|}{\|\mathbf{z}_c\|}$ is always positive due to the discriminant condition imposed by (B.14). We thus conclude that

$$\mathbf{z}^T \left(\frac{\beta_t}{\alpha_t} \bar{L} \otimes I_M + \mathcal{K}\mathcal{H} \right) \mathbf{z} > 0 \quad (\text{B.15})$$

for all $t \geq t_1$ and \mathbf{z} , such that $\|\mathbf{z}\| = 1$. Since the quadratic form in (B.15) is a continuous function on the compact unit circle, we may further conclude that

$$\inf_{\|\mathbf{z}\|=1} \mathbf{z}^T \left(\frac{\beta_t}{\alpha_t} \bar{L} \otimes I_M + \mathcal{K}\mathcal{H} \right) \mathbf{z} > c_2 > 0, \quad (\text{B.16})$$

for some positive constant c_2 , thus verifying the assertion in (B.10) for all $t \geq t_1$. To complete the proof of Proposition 5.5, choose any $0 < \varepsilon < c_2$. It then follows from (B.16) and continuity that for $t \geq t_1$ and arbitrary $\mathbf{z} \in \mathbb{R}^{NM}$,

$$\begin{aligned} \mathbf{z}^T \left(\beta_t \bar{L} \otimes I_M + \alpha_t \tilde{\mathcal{K}}\mathcal{H} \right) \mathbf{z} &\geq \alpha_t \|\mathbf{z}\|^2 \left[\inf_{\|\bar{\mathbf{z}}\|=1} \bar{\mathbf{z}}^T \left(\frac{\beta_t}{\alpha_t} \bar{L} \otimes I_M + \tilde{\mathcal{K}}\mathcal{H} \right) \bar{\mathbf{z}} \right] \\ &\geq (c_2 - \varepsilon) \alpha_t \|\mathbf{z}\|^2, \end{aligned}$$

thus verifying the assertion of Proposition 5.5 with $\varepsilon_K = \varepsilon$, $t_K = t_1$ and $c_K = c_2 - \varepsilon$. \square

Proof. [Proof of Proposition 5.6] By (B.13) in Proposition 5.5, there exists a constant $c_1 > 0$ such that for arbitrary $\mathbf{z} \in \mathbb{R}^{NM}$

$$\mathbf{z}^T \left(\frac{\beta_t}{\alpha_t} \bar{L} \otimes I_M + \mathcal{K}\mathcal{H} \right) \mathbf{z} \geq \left(\frac{\beta_t}{\alpha_t} \lambda_2(\bar{L}) - c_1 \right) \|\mathbf{z}_{c^\perp}\|^2 - 2c_1 \|\mathbf{z}_c\| \|\mathbf{z}_{c^\perp}\| + \|\mathbf{z}_c\|^2.$$

Hence for $\tilde{\mathcal{K}}$ satisfying (5.2), we have

$$\begin{aligned} \mathbf{z}^T \left(\frac{\beta_t}{\alpha_t} \bar{L} \otimes I_M + \tilde{\mathcal{K}}\mathcal{H} \right) \mathbf{z} &\geq \mathbf{z}^T \left(\frac{\beta_t}{\alpha_t} \bar{L} \otimes I_M + \mathcal{K}\mathcal{H} \right) \mathbf{z} - \varepsilon \|\mathbf{z}\|^2 \\ &= \left(\frac{\beta_t}{\alpha_t} \lambda_2(\bar{L}) - c_1 - \varepsilon \right) \|\mathbf{z}_{c^\perp}\|^2 - 2c_1 \|\mathbf{z}_c\| \|\mathbf{z}_{c^\perp}\| + (1 - \varepsilon) \|\mathbf{z}_c\|^2. \end{aligned}$$

Using the fact that $0 < \varepsilon < 1$, we have

$$\begin{aligned} \mathbf{z}^T \left(\frac{\beta_t}{\alpha_t} \bar{L} \otimes I_M + \tilde{\mathcal{K}}\mathcal{H} \right) \mathbf{z} &\geq \left(\frac{\beta_t}{2\alpha_t} \lambda_2(\bar{L}) + \left(\frac{\beta_t}{2\alpha_t} \lambda_2(\bar{L}) - c_1 - \varepsilon - \frac{c_1^2}{1-\varepsilon} \right) \right) \|\mathbf{z}_{C^\perp}\|^2 \\ &\quad + \left(\frac{c_1}{\sqrt{1-\varepsilon}} \|\mathbf{z}_{C^\perp}\| - \sqrt{1-\varepsilon} \|\mathbf{z}_C\| \right)^2. \end{aligned}$$

Since $\lambda_2(\bar{L}) > 0$ and $\beta_t/\alpha_t \rightarrow \infty$ as $t \rightarrow \infty$, there exists t_ε (large enough), such that,

$$\left(\frac{\beta_t}{2\alpha_t} \lambda_2(\bar{L}) - c_1 - \varepsilon - \frac{c_1^2}{1-\varepsilon} \right) \geq 0$$

for all $t \geq t_\varepsilon$. We may then conclude from (B.19) that

$$\mathbf{z}^T \left(\frac{\beta_t}{\alpha_t} \bar{L} \otimes I_M + \tilde{\mathcal{K}}\mathcal{H} \right) \mathbf{z} \geq \frac{\beta_t}{2\alpha_t} \lambda_2(\bar{L}) \|\mathbf{z}_{C^\perp}\|^2$$

and, hence

$$\mathbf{z}^T \left(\beta_t \bar{L} \otimes I_M + \alpha_t \tilde{\mathcal{K}}\mathcal{H} \right) \mathbf{z} \geq \frac{\lambda_2(\bar{L})}{2} \beta_t \|\mathbf{z}_{C^\perp}\|^2$$

for all $t \geq t_\varepsilon$, $\mathbf{z} \in \mathbb{R}^{NM}$ and \tilde{K} satisfying (5.2). This establishes the assertion. \square

Proof. [Proof of Lemma 5.7] Let the residual $\tilde{\mathbf{x}}_n(t) = \mathbf{x}_n(t) - \mathbf{x}_{\text{avg}}(t)$. Then arguments along the lines of (B.2)-(B.3) show that the process $\tilde{\mathbf{x}}_t = [\tilde{\mathbf{x}}_1^T(t), \dots, \tilde{\mathbf{x}}_N^T(t)]^T$ satisfies the recursion

$$\tilde{\mathbf{x}}_{t+1} = (I_{NM} - \beta_t L_t \otimes I_M) \tilde{\mathbf{x}}_t + \alpha_t \tilde{\mathbf{z}}_t,$$

where the process $\{\tilde{\mathbf{z}}_t\}$ is defined as

$$\tilde{\mathbf{z}}_t = \left(I_{NM} - \frac{1}{N} \mathbf{1}_N \otimes (\mathbf{1}_N \otimes I_M)^T \right) \mathcal{K}_t (\mathbf{y}_t - \mathcal{H}\mathbf{x}_t).$$

Since $\mathcal{K}_t \rightarrow \mathcal{K}$ as $t \rightarrow \infty$, the process $\{\mathbf{x}_t\}$ is bounded (Lemma 5.4), and the observation noise ζ_t satisfies **(A.5)**, there exist two \mathbb{R}_+ valued processes: 1) an \mathcal{F}_t -adapted $\{U_t\}$ satisfying $\sup_{t \geq 0} \|U_t\| < \infty$ a.s.; and (2) an i.i.d. $\{\mathcal{F}_{t+1}\}$ adapted $\{J_t\}$ independent of \mathcal{F}_t for each t and $\mathbb{E}_{\theta^*} [\|J_t\|^{2+\varepsilon_1}] < \infty$, such that

$$\|\tilde{\mathbf{z}}_t\| \leq U_t (1 + J_t).$$

Since $\tilde{\mathbf{x}}_t \in C^\perp$ for all t , by Lemma 4.4 there exists an $\{\mathcal{F}_{t+1}\}$ adapted \mathbb{R}_+ valued process $\{r_t\}$ with $0 \leq r_t \leq 1$ a.s. such that

$$\|(I_{NM} - \beta_t L_t \otimes I_M - \mathcal{P}_{NM}) \tilde{\mathbf{x}}_t\| \leq (1 - r_t) \|\tilde{\mathbf{x}}_t\|$$

for all t (large enough) and a constant $c_r > 0$ such that for all t

$$\mathbb{E}_{\theta^*} [r_t \mid \mathcal{F}_t] \geq \frac{c_r}{(t+1)^{\tau_2}} \quad \text{a.s.}$$

From the above development we conclude that

$$\|\tilde{\mathbf{x}}_{t+1}\| \leq (1 - r_t) \|\tilde{\mathbf{x}}_t\| + \alpha_t U_t (1 + J_t) \tag{B.20}$$

for all t (large enough). The recursion (B.20) clearly falls under the purview of Lemma 4.3, and we have the assertion

$$\mathbb{P}_{\theta^*} \left(\lim_{t \rightarrow \infty} (t+1)^{\tau_0} \tilde{\mathbf{x}}_t = 0 \right) = 1$$

for all $\tau_0 \in \left[0, \tau_1 - \tau_2 - \frac{1}{2+\varepsilon_1} \right)$. This establishes the claim. \square

Proof of Lemma 5.9

We will use the following approximation result from [42] in the proof.

PROPOSITION B.1 (Lemma 4.3 in [42]). *Let $\{b_t\}$ be a scalar sequence satisfying*

$$b_{t+1} \leq \left(1 - \frac{c}{t+1}\right) b_t + d_t(t+1)^{-\tau}$$

where $c > \tau$, $\tau > 0$, and the sequence $\{d_t\}$ is summable. Then $\limsup_{t \rightarrow \infty} (t+1)^\tau b_t < \infty$.

The following generalized convergence criterion of dependent stochastic sequences will also be useful.

PROPOSITION B.2 (Lemma 10 in [43]). *Let $\{\bar{J}_t\}$ be an \mathbb{R} valued $\{\mathcal{F}_{t+1}\}$ adapted process such that $\mathbb{E}[\bar{J}_t | \mathcal{F}_t] = 0$ a.s. for each $t \geq 1$. Then the sum $\sum_{t \geq 0} \bar{J}_t$ exists and is finite a.s. on the set where $\sum_{t \geq 0} \mathbb{E}[\bar{J}_t^2 | \mathcal{F}_t]$ is finite.*

Proof. [Proof of Lemma 5.9] For each $\delta > 0$ recall the construction in (5.8)-(5.10). Clearly, it suffices by the arguments in Lemma 5.8 to establish the required convergence rate claim for each of the processes $\{\mathbf{z}_t^\delta\}$.

Let $\bar{\tau} \in [0, 1/2)$ be such that

$$\mathbb{P}_{\theta^*} \left(\lim_{t \rightarrow \infty} (t+1)^{\bar{\tau}} \|\mathbf{z}_t^\delta\| = 0 \right) = 1$$

for all n . Such a $\bar{\tau}$ always exists by Lemma 5.8. We now show that there exists τ such that $\bar{\tau} < \tau < 1/2$ for which the claim holds. To this end, choose $\tilde{\tau} \in (\tau, 1/2)$ and let $\mu = 1/2(\bar{\tau} + \tilde{\tau})$. For each $\delta > 0$ recall the construction in (5.8)-(5.10) and the \mathcal{F}_t -adapted process $\{\mathbf{z}_t^\delta\}$ satisfies

$$\begin{aligned} \|\mathbf{z}_{t+1}^\delta\|^2 &\leq \|I_M - \alpha_t \Gamma_t^\delta\|^2 \|\mathbf{z}_t^\delta\|^2 + \alpha_t^2 \|U_t^\delta\|^2 + \alpha_t^2 \|J_t^\delta\|^2 + 2\alpha_t (\mathbf{z}_t^\delta)^T (I_M - \alpha_t \Gamma_t^\delta) J_t^\delta \\ &\quad + 2\alpha_t \|U_t^\delta\| \left(\|I_M - \alpha_t \Gamma_t^\delta\| \|\mathbf{z}_t^\delta\| + \alpha_t \|J_t^\delta\| \right). \end{aligned}$$

Since $\tau_1 > \tau_2 + 1/(2 + \varepsilon_1) + 1/2$, by Lemma 5.7 and (5.9) the process $\{U_t^\delta\}$ may be chosen such that²

$$\|U_t^\delta\| = o\left((t+1)^{-1/2}\right). \quad (\text{B.22})$$

Since $\|\mathbf{z}_t^\delta\| = o\left((t+1)^{-\bar{\tau}}\right)$ (by hypothesis), we obtain

$$2\alpha_t \|U_t^\delta\| \|I_M - \alpha_t \Gamma_t^\delta\| \|\mathbf{z}_t^\delta\| = o\left((t+1)^{-3/2-\bar{\tau}}\right).$$

The existence of the second moment of the observation noise process and the boundedness of $\{\mathcal{K}_t^\delta\}$ imply

$$\mathbb{P}_{\theta^*} \left(\lim_{t \rightarrow \infty} (t+1)^{-1/2-\varepsilon} \|J_t^\delta\| = 0 \right) = 1 \quad (\text{B.23})$$

for each $\varepsilon > 0$ and, hence

$$2\alpha_t^2 \|U_t^\delta\| \|J_t^\delta\| = o\left((t+1)^{-3/2-\bar{\tau}}\right).$$

Since $2\mu = \bar{\tau} + \tilde{\tau}$ and $\tilde{\tau} < 1/2$, by (B.23) we note that

$$\sum_{t \geq 0} (t+1)^{2\mu} \alpha_t \|U_t^\delta\| \|I_M - \alpha_t \Gamma_t^\delta\| \|\mathbf{z}_t^\delta\| < \infty.$$

Similarly we have

$$\sum_{t \geq 0} (t+1)^{2\mu} \alpha_t^2 \|U_t^\delta\| \|J_t^\delta\| < \infty, \quad \sum_{t \geq 0} (t+1)^{2\mu} \alpha_t^2 \|U_t^\delta\|^2 < \infty.$$

²For \mathbb{R}_+ valued sequences $\{f_t\}$ and $\{g_t\}$ the notation $f_t = o(g_t)$ means that $f_t/g_t \rightarrow 0$ as $t \rightarrow \infty$. For stochastic sequences the $o(\cdot)$ is to be interpreted a.s. or pathwise.

Now consider the terms $\alpha_t^2 \|J_t^\delta\|^2$. Since the second moment of the observation noise process exists, $\{\mathcal{K}_t^\delta\}$ is uniformly bounded and $2\mu < 1$, it can be shown that

$$\sum_{t \geq 0} (t+1)^{2\mu} \alpha_t^2 \|J_t^\delta\|^2 < \infty.$$

Now let $\{W_t^\delta\}$ denote the \mathcal{F}_{t+1} sequence given by

$$W_t^\delta = \alpha_t \left(\mathbf{z}_t^\delta \right)^T \left(I_M - \alpha_t \Gamma_t^\delta \right) J_t^\delta.$$

We note that $\mathbb{E}_{\theta^*}[W_t^\delta | \mathcal{F}_t] = 0$ for all t and (at least for t large) we have $\mathbb{E}_{\theta^*}[(W_t^\delta)^2 | \mathcal{F}_t] \leq \alpha_t^2 \|\mathbf{z}_t^\delta\|^2 \|J_t^\delta\|^2$. Since the second moment of the observation noise process exists and $\{\mathcal{K}_t^\delta\}$ is uniformly bounded, we obtain

$$\mathbb{E}_{\theta^*} \left[\left(W_t^\delta \right)^2 \mid \mathcal{F}_t \right] = o \left((t+1)^{-2-2\bar{\tau}} \right).$$

Hence

$$\mathbb{E}_{\theta^*} \left[(t+1)^{4\mu} \left(W_t^\delta \right)^2 \mid \mathcal{F}_t \right] = o \left((t+1)^{-2-2\bar{\tau}+4\mu} \right) = o \left((t+1)^{-2+2\bar{\tau}} \right). \quad (\text{B.24})$$

Since $2\bar{\tau} < 1$, the sequence on the left hand side of (B.24) is summable and by Proposition B.2 we conclude that $\sum_{t \geq 0} (t+1)^{2\mu} W_t^\delta$ exists and is finite. Since $\Gamma_t^\delta \rightarrow I_M$ uniformly and $\alpha_t \rightarrow 0$ as $t \rightarrow \infty$, we have

$$\left\| I_M - \alpha_t \Gamma_t^\delta \right\|^2 \leq (1 - a(t+1)^{-1}) \quad (\text{B.25})$$

for all t large enough. Thus (eventually) we have from (B.21)

$$\left\| \mathbf{z}_{t+1}^\delta \right\|^2 \leq (1 - a(t+1)^{-1}) \left\| \mathbf{z}_t^\delta \right\|^2 + d_t (t+1)^{-2\mu}$$

where the term $d_t (t+1)^{-2\mu}$ corresponds to all the residuals. Moreover by (B.22)-(B.25) the limit $\lim_{t \rightarrow \infty} \sum_{s=0}^t d_s$ exists and is finite. Since $a \geq 1 > 2\mu$, an immediate application of Proposition B.1 yields

$$\limsup_{t \rightarrow \infty} (t+1)^{2\mu} \left\| \mathbf{z}_t^\delta \right\|^2 < \infty \quad \text{a.s.}$$

Hence, there exists τ with $\bar{\tau} < \tau < \mu$, such that $(t+1)^\tau \left\| \mathbf{z}_t^\delta \right\| \rightarrow 0$ a.s. as $t \rightarrow \infty$. Since the above holds for all $\delta > 0$, we conclude that $(t+1)^\tau \left\| \mathbf{z}_t \right\| \rightarrow 0$ a.s. as $t \rightarrow \infty$. Thus, for every $\bar{\tau}$ for which the convergence in (5.16) holds there exists $\tau \in (\bar{\tau}, 1/2)$ for which the convergence continues to hold. Hence, by induction we conclude that the required convergence holds for all $\tau \in [0, 1/2)$. \square

REFERENCES

- [1] R. Olfati-Saber, J. A. Fax, and R. M. Murray. Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 95(1):215–233, Jan. 2007.
- [2] A. G. Dimakis, S. Kar, J.M.F. Moura, M. G. Rabbat, and A. Scaglione. Gossip algorithms for distributed signal processing. *Proceedings of the IEEE*, 98(11):1847–1864, Nov. 2010.
- [3] T. Chung, V. Gupta, J. Burdick, and R. Murray. On a decentralized active sensing strategy using mobile sensor platforms in a network. In *43rd IEEE Conference on Decision and Control*, volume 2, pages 1914–1919, Paradise Island, Bahamas, Dec. 2004.
- [4] L. Xiao, S. Boyd, and S. Lall. A scheme for robust distributed sensor fusion based on average consensus. In *Proceedings of the International Conference on Information Processing in Sensor Networks*, pages 63–70, Los Angeles, CA, April 2005.
- [5] I. D. Schizas, A. Ribeiro, and G. B. Giannakis. Consensus in ad hoc WSNs with noisy links - Part I: Distributed estimation of deterministic signals. *IEEE Transactions on Signal Processing*, 56(1):350–364, Jan. 2008.

- [6] S. Kar, J. M. F. Moura, and K. Ramanan. Distributed parameter estimation in sensor networks: nonlinear observation models and imperfect communication. *IEEE Transactions on Information Theory*, 58(8):3575–3605, Jun. 2012.
- [7] A. Jadbabaie, A. Tahbaz-Salehi, and A. Sandroni. Non-Bayesian social learning. PIER Working paper 2010-005, University of Pennsylvania, Philadelphia, PA, Feb. 2010.
- [8] D. Jakovetic, J. Xavier, and J.M.F. Moura. Weight optimization for consensus algorithms with correlated switching topology. *IEEE Transactions on Signal Processing*, 58(7):3788–3801, July 2010.
- [9] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri. Distributed Kalman filtering using consensus strategies. In *Proceedings of the 46th IEEE Conference on Decision and Control*, pages 5486–5491, 2007.
- [10] U. A. Khan and J. M. F. Moura. Distributing the Kalman filter for large-scale systems. *IEEE Transactions on Signal Processing*, 56(1)(10):4919–4935, Oct. 2008.
- [11] C.G. Lopes and A.H. Sayed. Diffusion least-mean squares over adaptive networks: Formulation and performance analysis. *IEEE Transactions on Signal Processing*, 56(7):3122–3136, Jul. 2008.
- [12] I.D. Schizas, G. Mateos, and G.B. Giannakis. Stability analysis of the consensus-based distributed LMS algorithm. In *Proceedings of the 33rd International Conference on Acoustics, Speech, and Signal Processing*, pages 3289–3292, Las Vegas, Nevada, USA, Apr. 1-4 2008.
- [13] R. Olfati-Saber. Kalman-consensus filter: optimality, stability, and performance. In *48th IEEE Conference on Decision and Control*, pages 7036–7042, Shanghai, China, Dec. 2009.
- [14] S. S. Ram, A. Nedic, and V. V. Veeravalli. Incremental stochastic subgradient algorithms for convex optimization. *SIAM Journal on Optimization*, 20(2):691–717, Jun. 2009.
- [15] I. Lobel and A. Ozdaglar. Distributed subgradient methods for convex optimization over random networks. *IEEE Transactions on Automatic Control*, 56(6):1291–1306, Jun. 2011.
- [16] G. Yin, Y. Sun, and L.Y. Wang. Asymptotic properties of consensus-type algorithms for networked systems with regime-switching topologies. *Automatica*, 47(7):1366–1378, Jul. 2011.
- [17] D. Jakovetic, J. Xavier, and J.M.F. Moura. Cooperative convex optimization in networked systems: augmented Lagrangian algorithms with directed gossip communication. *IEEE Transactions on Signal Processing*, 59(8):3889–3902, Aug. 2011.
- [18] T. L. Lai and C. Z. Wei. Asymptotically efficient self-tuning regulators. *SIAM J. Control and Optimization*, 25(2):466–481, Mar. 1987.
- [19] T. L. Lai. Asymptotic properties of nonlinear least squares estimates in stochastic regression models. *The Annals of Statistics*, 2(4):1917–1930, 1994.
- [20] S. B. Gelfand and S. K. Mitter. Recursive stochastic algorithms for global optimization in \mathbb{R}^d . *SIAM J. Control Optim.*, 29(5):999–1018, Sept. 1991.
- [21] V. S. Borkar. *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge University Press, Cambridge, UK, 2008.
- [22] G. Yin and Q. Zhang. *Discrete-time Markov Chains: Two-time-scale Methods and Applications*. Springer, New York, NY, 2005.
- [23] V. Krishnamurthy, K. Topley, and G. Yin. Consensus formation in a two-time-scale Markovian system. *SIAM Journal on Multiscale Modeling and Simulation*, 7(4):1898–1927, Jul. 2009.
- [24] M.B. Nevel’son and R.Z. Has’minskii. *Stochastic Approximation and Recursive Estimation*. American Mathematical Society, Providence, Rhode Island, 1973.
- [25] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Transactions on Automatic Control*, 31(9):803–812, Sept. 1986.
- [26] D. Bertsekas, J.N. Tsitsiklis, and M. Athans. Convergence theories of distributed iterative processes: A survey. *Technical Report for Information and Decision Systems, Massachusetts Inst. of Technology, Cambridge, MA*, 1984.
- [27] H.J. Kushner and G. Yin. Asymptotic properties of distributed and communicating stochastic approximation algorithms. *SIAM J. Control and Optimization*, 25(5):1266–1290, Sept. 1987.
- [28] S.S. Stankovic, M.S. Stankovic, and D.M. Stipanovic. Decentralized parameter estimation by consensus based stochastic approximation. In *46th IEEE Conference on Decision and Control*, pages 1535–1540, New Orleans, LA, USA, 12-14 Dec. 2007.
- [29] S. Kar and J. M. F. Moura. Distributed consensus algorithms in sensor networks with imperfect communication: link failures and channel noise. *IEEE Transactions on Signal Processing*, 57(1):355–369, Jan. 2009.
- [30] T. Li, M. Fu, L. Xie, and J.-F. Zhang. Distributed consensus with limited communication data

- rate. *IEEE Transactions on Automatic Control*, 56(2):279–292, Feb. 2011.
- [31] M. Huang and J.H. Manton. Stochastic approximation for consensus seeking: mean square and almost sure convergence. In *Proceedings of the 46th IEEE Conference on Decision and Control*, New Orleans, LA, USA, Dec. 12-14 2007.
- [32] F. R. K. Chung. *Spectral Graph Theory*. Providence, RI : American Mathematical Society, 1997.
- [33] S. Kar and J.M.F. Moura. Convergence rate analysis of distributed gossip (linear parameter) estimation: fundamental limits and tradeoffs. *IEEE Journal of Selected Topics in Signal Processing Signal Processing in Gossiping Algorithms Design and Applications*, 5(4):674–690, Aug. 2011.
- [34] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Transactions on Information Theory*, 52(6):2508 – 2530, Jun. 2006.
- [35] S. Tatikonda and S. Mitter. Control under communication constraints. *IEEE Transactions on Automatic Control*, 49(7):1056 – 1068, Jul. 2004.
- [36] A. Matveev and A. Savkin. The problem of state estimation via asynchronous communication channels with irregular transmission times. *IEEE Transactions on Automatic Control*, 48(4):670–676, Apr. 2006.
- [37] K. Li and J. Baillieul. Robust and efficient quantization and coding for control of multidimensional linear systems under data rate constraints. *International Journal of Robust and Nonlinear Control Special Issue: Communicating-Agent Networks*, 17(10-11):898–920, Jul. 2007.
- [38] B.T. Polyak. *Introduction to Optimization*. Optimization Software, Inc., New York, NY, 1987.
- [39] P. R. Halmos. *Measure Theory*. Springer-Verlag, New York, 1974.
- [40] V. Fabian. On asymptotic normality in stochastic approximation. *The Annals of Mathematical Statistics*, 39(4):1327–1332, Aug. 1968.
- [41] P. Billingsley. *Convergence of Probability Measures*. John Wiley and Sons, Inc., 1999.
- [42] V. Fabian. Stochastic approximation of minima with improved asymptotic speed. *The Annals of Mathematical Statistics*, 37(1):191–200, Feb. 1967.
- [43] L. E. Dubins and D. A. Freedman. A sharper form of the Borel-Cantelli lemma and the strong law. *The Annals of Mathematical Statistics*, 36(3):800–807, Jun. 1965.